THIRD EDITION

Valve Amplifiers

MORGAN JONES

VISIT…

Valve Amplifiers

This page intentionally left blank

Valve Amplifiers

Third edition

Morgan Jones

AMSTERDAM BOSTON HEIDELBERG LONDON NEW YORK OXFORD
PARIS SAN DIEGO SAN FRANCISCO SINGAPORE SYDNEY TOKYO

Newnes
An imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP
200 Wheeler Road, Burlington, MA 01803

First published 1995
Second edition 1999
Third edition 2003

Copyright © 1995, 1999, 2003, Morgan Jones. All rights reserved.

The right of Morgan Jones to be identified as the author of
this work has been asserted in accordance with the
Copyright, Designs and Patents Act 1988

No part of this publication may be reproduced in
any material form (including photocopying or storing in any medium by
electronic means and whether or not transiently or incidentally to some
other use of this publication) without the written permission of the
copyright holder except in accordance with the provisions of the Copyright,
Designs and Patents Act 1988 or under the terms of a licence issued by the
Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London,
England W1T 4LP. Applications for the copyright holder's written
permission to reproduce any part of this publication should be addressed
to the publisher

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0 7506 56948

All photographs (including front cover) by Morgan Jones

For information on all Newnes publications
visit our website at www.newnespress.com

Typeset by Integra Software Services Pvt. Ltd, Pondicherry, India
www.integra-india.com
Printed and bound in Great Britain

Contents

Preface ix

Dedication xi

Acknowledgements xiii

1 Circuit analysis **1**
Mathematical symbols 1
Electrons and definitions 2
Potential dividers 15
Alternating current 23
Active devices 44
Silicon diodes 45
Bipolar Junction Transistors (BJTs) 48
General observations on BJTs 54
Feedback 54
The operational amplifier 57
References 63
Further reading 63

2 Basic building blocks **64**
The common cathode triode amplifier 64
The tetrode 86
The beam tetrode and the pentode 88
The cascode 94
The cathode follower 100
Sources and sinks: definitions 104

The common cathode amplifier as a constant current sink 105
The cathode follower with active load 109
The White cathode follower 110
The μ-follower 115
The Shunt Regulated Push–Pull (SRPP) amplifier 120
The β-follower 125
The differential pair 127
Semiconductor constant current sinks 133
References 140
Further reading 141

3 Distortion 143
Defining distortion 143
Digital concepts 151
The Fast Fourier Transform (FFT) 154
Designing for low distortion 157
AC operating point 157
DC operating point 161
Distortion reduction by parameter restriction 166
Distortion reduction by cancellation 169
DC bias problems 170
Individual valve choice 178
Coupling from one stage to the next 189
References 200
Further reading 201

4 Component technology 202
Resistors 202
General considerations on choosing resistors 211
Capacitors 211
Different types of capacitors 214
General considerations in choosing capacitors 225
Magnetic components 229
Inductors 229
Transformers 235
Why should I use a transformer? 247
General considerations in choosing transformers 248
Uses and abuses of audio transformers 253
Thermionic valves 255
Individual elements of the valve structure 259

References 289
Further reading 290

5 Power supplies 292
The major blocks 292
Rectification and smoothing 293
Regulators 328
A practical design 351
Specification of the power supply 351
A better power supply 365
References 376
Further reading 376

6 The power amplifier 377
The output stage 377
Classes of amplifiers 383
The push–pull output stage and the output transformer 386
Output Transformer-Less (OTL) amplifiers 393
The entire amplifier 394
The driver stage 394
The phase splitter 396
The input stage 407
Stability 408
Classic power amplifiers 412
New designs 423
Single-ended madness 424
The 'Scrapbox Challenge' single-ended amplifier 424
HT choke suitability 431
Push–pull sanity 444
The 'Bevois Valley' amplifier 445
The author's prototypes 453
The traditional approach to power >10 W 455
Driving higher power output stages 458
The 'Crystal Palace' amplifier 460
'Daughter of Beast' 483
References 489
Further reading 490

7 The pre-amplifier 491
The line stage 491

The volume control 500
Input selectors 513
The RIAA disc stage 516
Noise and input capacitance of the input stage 528
Strays and implementing RIAA 537
Basic pre-amplifier compromises 547
A balanced pre-amplifier 548
The perfect line stage 555
The EC8010 RIAA stage 557
Achieving the RIAA equalization 563
Practical RIAA considerations 570
A line stage 572
References 577
Further reading 578

Appendix **579**
Valve data 579
Square wave sag and low frequency $f_{-3\,dB}$ 585
Playing 78s 587
Equalization 589
Standard component values 593
Resistor colour code 594
Sallen & Key filters for active crossovers 596
UK sources of components 600
Bargains and dealing directly 601
References 602

Index 603

Preface

Thirty years ago the author bought his first valve amplifier; it cost him £3, and represented many weeks' pocket money. Whilst his pocket money has increased, so have his aspirations, and the DIY need was born.

Although there were many sources of information on circuit design, the electronics works gave scant regard to audio design, whilst the Hi-Fi books barely scratched the surface of the theory. The author therefore spent much time in libraries trying to link this information together to form a basis for audio design. This book is the result of those years of effort and aims to present thermionic theory in an accessible form without getting too bogged down in maths. Primarily, it is a book for practical people armed with a cheap scientific calculator, a power drill, and a (temperature-controlled) soldering iron.

The author started a BSc in Acoustical Engineering, but left after a year to join BBC Engineering as a Technical Assistant, where he received excellent tuition in electronics and rose to the giddy heights of a Senior Engineer before being made redundant by BBC cuts. More recently, he has served time in Higher Education, and although developing the UK's first BSc (Hons) Media Technology course and watching students blossom into graduates with successful careers was immensely satisfying, education is achieved by class contact – not by committees and paper chases.

Early on, he became a member of the Audio Engineering Society, and has designed and constructed many valve pre-amplifiers and power amplifiers, loudspeakers, pick-up arms, and a pair of electrostatic headphones.

It is now almost a decade since work began on the 1st edition of *Valve Amplifiers*, yet much has changed in this obsolete technology since then.

The relentless infestation of homes by computers has meant that test and measurement has become much cheaper, either because it directly uses the

processing power of a computer, or because it borrows from the technology needed to make them. Thus, the Fast Fourier Transform has become a tool for all to use, from industrial designer to keen amateur – enabling detailed spectrum analysis that was the province of world class companies only 15 years ago. As a happy consequence, this edition benefits from detailed measurements limited primarily by the author's time.

The spread of Internet trading has made the market for valves truly global. Exotica such as Loctals, European 'Special Quality' valves, and final generation Soviet bloc valves are now all readily available worldwide to any Luddite with the patience to access the Internet – we no longer need to be constrained to conservative choices of traditional audio valves.

Paradoxically, although digital electronics has improved the supply of valves, it is killing analogue components. Capacitors are the worst affected. Silvered-mica capacitors are all but extinct, and polystyrene is now an endangered species. Controls everywhere have succumbed to the ubiquitous digital encoder, so mechanical switch ranges have contracted, and potentiometers face a similar Darwinian fate.

Despite, or perhaps *because* of these problems, valves and vinyl have become design icons, both in television adverts, and the bits in between. A 2002 UK advertisement flashed an 845 as part of its spiel for a completely unrelated product.

Eagle-eyed readers will note that V_{be} is now quoted as 0.7 V rather than 0.6 V. BBC-designed equipment ran hot in its bay, so 0.6 V was appropriate, but Higher Education cannot afford such high ambient temperatures, so 0.7 V is more common.

Note for the CD generation: Pick-up arms are exquisitely engineered mechanical contrivances that support a lump of rock on the end of a scaffold pole being scraped by the wiggly groove in a 300 mm (flattish) vinyl disc known as an LP. The resulting sound is often very good.

Dedication

The author would like to dedicate this book to the dwindling band of BBC engineers, particularly at BBC Southampton, and also to those at BBC Wood Norton, of which he has many colourful memories.

This page intentionally left blank

Acknowledgements

As in the first edition, the book, if not the author, has benefited from unintentional encouragement.

Special thanks are due to Paul Leclercq, who undertook the onerous task of proofreading. Without his eagle eye, more infinitives would have been split, and fewer of the more obscure valves could have been obtained for the distortion test programme detailed in Chapter 3.

Many individuals helped research material, but significant contributions were made by: Dave Mansell, Paul Leclercq, Gary Longrie, Roger Lownsborough, and Brian Terrell.

Likewise, many commercial enterprises were most helpful with information or loaned equipment, and their names have been mentioned in the text, but particular credit is due to Brian Sowter of Sowter Transformers for his patience in listening to the author's bizarre requests.

Finally, the author would like to thank those readers who took the time and trouble to breach the publishing citadel and give the author hugely useful feedback.

This page intentionally left blank

1

Circuit analysis

In order to look at the interesting business of designing and building valve amplifiers, we will need some knowledge of electronics fundamentals. Unfortunately, fundamentals are not terribly interesting, and to cover them fully would consume the entire book. Ruthless pruning has therefore been necessary to condense what is needed into one chapter.

It is thus with deep sorrow that the author has had to forsake complex numbers and vectors, whilst the omission of differential calculus is a particularly poignant loss. All that is left is ordinary algebra, and although there are *lots* of equations, they are timid, miserable creatures, and quite defenceless.

If you are comfortable with basic electronic terms and techniques, then please feel free to skip directly to Chapter 2, where valves appear.

Mathematical symbols

Unavoidably, a number of mathematical symbols are used, some of which you may have forgotten, or perhaps not previously met:

$a \equiv b$ a is totally equivalent to b
$a = b$ a equals b
$a \approx b$ a is approximately equal to b
$a \propto b$ a is proportional to b
$a \neq b$ a is not equal to b
$a > b$ a is greater than b
$a < b$ a is less than b
$a \geq b$ a is greater than, or equal to, b
$a \leq b$ a is less than, or equal to, b

As with the $=$ and $\neq$ symbols, the four preceding symbols can have a slash through them to negate their meaning ($a \not< b$, a is not less than b).

$\sqrt{a}$	the number which when multiplied by itself is equal to a (square root)
a^n	a multiplied by itself n times. $a^4 = a \times a \times a \times a$ (a to the power n)
$\pm$	plus or minus
∞	infinity
$^\circ$	degrees, either of temperature (°C), or of an angle (360° in a circle)
$\|\|$	parallel, either parallel lines, or an electrical parallel connection

Electrons and definitions

Electrons are *charged* particles. Charged objects are attracted to other charged particles, or objects. A practical demonstration of this is to take a balloon and rub it briskly against a jumper, then place the rubbed face against a wall. Let go. The balloon remains stuck to the wall. This is because we have charged the balloon, and so there is an attractive force between it and the wall. (Although the wall was initially uncharged, placing the balloon on the wall induced a charge.)

Charged objects come in two flavours: negative and positive. Unlike charges attract and like charges repel. Electrons are negative and, if we could find some, positrons would be positive.

If we don't have ready access to positrons, how can we have a positively charged object? Suppose we had an object that was negatively charged, because it had 2000 electrons clustered on its surface. If we had another, similar, object that only had 1000 electrons on its surface, then we would say that the first object was more negatively charged than the second, but as we can't count how many electrons we have, we might just as easily have said that the second object was more positively charged than the first. It's just a matter of which way you look at it.

To charge our balloon, we had to do some work, and use energy. We had to overcome friction when rubbing the balloon against the woollen jumper. In the process, electrons were moved from one surface to the other. Therefore one object (the balloon) has acquired an excess of electrons and is negatively charged, whilst the woollen jumper has lost the same number of electrons and is positively charged.

The balloon would therefore stick to the jumper. Or would it? Certainly it will be attracted to the jumper, but what happens when we place the two in contact? The balloon does not stick. This is because the fibres of the jumper were able to touch all of the charged area on the balloon, and the electrons

were so attracted to the jumper that they moved back onto the jumper, thus neutralizing the charge.

At this point we can discard vague talk of balloons and jumpers because we have just observed electron flow.

An electron is very small, and doesn't have much of a charge, so we need a more practical unit for defining charge. That practical unit is the *coulomb* (C). We could now say that 1 coulomb of charge had flowed between one point and another, which would be equivalent to saying that approximately 6 180 000 000 000 000 000 electrons had passed, but much handier.

Simply being able to say that a large number of electrons had flowed past a given point is not in itself very helpful. We might say that a billion cars have travelled down a particular section of motorway since it was built, but if you were planning a journey down that motorway, you would want to know the flow of cars *per hour* through that section.

Similarly in electronics, we are not concerned with the total flow of electrons since the dawn of time, but we do want to know about electron flow at any given instant. Thus, we could define the flow as the number of coulombs of charge that flowed past a point in one second. This is still rather long winded and we will abbreviate yet further.

We will call the flow of electrons a *current*, and, as the coulomb/second is unwieldy, it will be redefined as a new unit, the *ampere* (A). Because the ampere is such a useful unit, the definition linking current and charge is usually stated in the following form:

One coulomb is the charge moved by one ampere flowing for one second.

$$\text{charge (coulombs)} = \text{current (amperes)} \times \text{time (seconds)}$$

This is still rather unwieldy, so symbols are assigned to the various units: charge has the symbol Q, current I, and time t:

$$Q = It$$

This is a very useful equation, and we will meet it again when we look at capacitors (which store charge).

Meanwhile, current has been flowing, but why did it flow? If we are going to move electrons from one place to another, we need a force to cause this

movement. This force is known as the *Electro-Motive Force* (EMF). Current continues to flow whilst this force is applied, and flows from a higher potential to a lower potential.

If two points are at the same potential, no current can flow between them. What is important is the *potential difference* (pd).

A potential difference causes a current to flow between two points. As this is a new property, we need a unit, a symbol and a definition to describe it. We mentioned work being done in charging the balloon, and in its very precise and physical sense, this is how we can define potential difference, but first, we must define *work*:

One joule of work is done if a force of one newton moves one metre from its point of application.

This very physical interpretation of work can be understood easily once we realize that it means that one joule of work would be done by moving one kilogramme a distance of one metre in one second. Since charge is directly related to the mass of electrons moved, the physical definition of work can be modified to define the force that causes the movement of charge.

Unsurprisingly, because it causes the motion of electrons, the force is called the Electro-Motive Force (EMF), and it is measured in *volts*:

If one joule of work is done moving one coulomb of charge, then the system is said to have a potential difference of one volt (V).

$$\text{work done (joules)} = \text{charge (coulombs)} \times \text{potential difference (volts)}$$

$$W = QV$$

The concept of work is important because work can only be done by the expenditure of energy, which is therefore also expressed in joules:

$$\text{work done (joules)} = \text{energy expended (joules)}$$

$$W = E$$

In our specialized sense, doing work means moving charge (electrons) to make currents flow.

Batteries and lightbulbs

If we want to make a current flow, we need a *circuit*. A circuit is exactly that; a loop or path through which a current can flow, from its starting point all the way round the circuit, to *return* to its starting point. Break the circuit and the current ceases to flow.

The simplest circuit that we might imagine is a battery connected to a lightbulb via a switch. We *open* the switch to stop the current flow (open circuit), and close it to light the bulb. Meanwhile our helpful friend (who has been watching all this) leans over and drops a thick piece of copper across the battery terminals, causing a *short* circuit.

The bulb goes out. Why?

Ohm's law

To answer the last question we need some property that defines how much current flows. That property is *resistance*, so we will need another definition, units, and a symbol:

If a potential difference of one volt is applied across a resistance, resulting in a current of one ampere, then the resistance has a value of one ohm (Ω).

$$\text{potential difference (volts)} = \text{current (amperes)} \times \text{resistance (ohms)}$$

$$V = IR$$

This is actually a simplified statement of Ohm's law, rather than a strict definition of resistance, but we don't need to worry too much about that.

We can rearrange the previous equation to make I or R the subject:

$$I = \frac{V}{R}$$

$$R = \frac{V}{I}$$

These are incredibly powerful equations, and should be committed to memory.

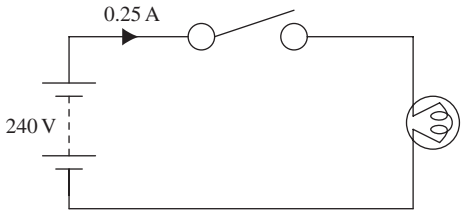


Fig. 1.1 *Use of Ohm's law to determine the resistance of a hot lightbulb*

The circuit in Fig. 1.1 is switched on, and a current of 0.25 A flows. What is the resistance of the bulb?

$$R = \frac{V}{I} = \frac{240}{0.25} = 960\,\Omega$$

Now this might seem like a trivial example, since we could easily have measured the resistance of the bulb to 4½ significant figures using our shiny, new, digital multimeter. But could we? The hot resistance of a lightbulb is very different from its cold resistance; in the example above, the cold resistance was 80 Ω.

We could now work the other way, and ask what current would flow through an 80 Ω resistor connected to 240 V?

$$I = \frac{V}{R} = \frac{240}{80} = 3\,\text{A}$$

Incidentally, this is why lightbulbs are most likely to fail at switch-on. The high initial current that flows before the filament has warmed up and increased its resistance, stresses the weakest parts of the filament, they become so hot that they vaporize, and the bulb blows.

Power

In the previous example, we looked at a lightbulb, and rated it by the current that flowed through it when connected to a 240 V battery. But we all know that lightbulbs are rated in *watts*, so there must be some connection between the two:

One watt (W) of power is expended if one joule of work is done in one second.

$$\text{power (watts)} = \frac{\text{work done (joules)}}{\text{time taken (seconds)}}$$

$$P = \frac{W}{t}$$

This may not seem to be the most useful of definitions, and indeed it isn't, but by combining it with some earlier equations:

$$W = QV$$

So:

$$P = \frac{QV}{t}$$

But:

$$Q = It$$

So:

$$P = \frac{IVt}{t}$$

We obtain:

$$P = IV$$

This is a fundamental equation of equal importance to Ohm's law. Substituting the Ohm's law equations into this yields:

$$P = \frac{V^2}{R}$$
$$= I^2R$$

We can now use these equations to calculate the power rating of our lightbulb. Since it drew 0.25 A when fed from 240 V, and had a hot resistance of 960 Ω, we can use any of the three equations.

Using:

$$\begin{aligned} P &= \frac{V^2}{R} \\ &= \frac{240^2}{960} \\ &= 60\,\mathrm{W} \end{aligned}$$

It will probably not have escaped your notice that this lightbulb looks suspiciously like an AC mains lightbulb, and that the battery was rather large. We will return to this later.

Kirchhoff's laws

There are two of these: a current law and a voltage law. They are both very simple and at the same time very powerful.

The current law states:

The algebraic sum of the currents flowing into, and out of, a node is equal to zero.

$$0 = I_1 + I_2 + I_3 + \cdots$$

What it says in a more relaxed form is that what goes in, comes out. If we have 10 A going into a node, or junction, then that current must also leave that junction – it might not all come out on one wire, but it will all come out. A sort of conservation of current, if you like. See Fig. 1.2.

Current flowing into the node: $I_1 = 10\,\mathrm{A}$
Currents leaving the node: $I_2 = 4\,\mathrm{A}$
$I_3 = 6\,\mathrm{A}$
Total current leaving the node: $I_{\text{total}} = 10\,\mathrm{A}$

From the point of view of the node, the currents leaving the node are flowing in the opposite direction to the current flowing into the

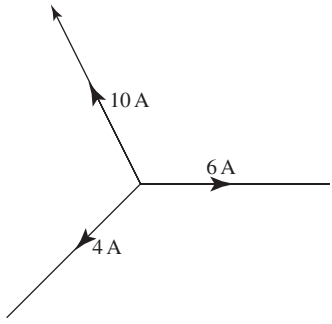


Fig. 1.2 *Currents at a node (Kirchhoff's current law)*

node, so we *must* give them a minus sign before plugging them into the equation.

$$\begin{aligned} 0 &= I_1 + I_2 + I_3 \\ &= 10\,\text{A} + (-4\,\text{A}) + (-6\,\text{A}) \\ &= 10 - 4 - 6 \end{aligned}$$

This may have seemed pedantic, since it was obvious from the diagram that the incoming currents equalled the outgoing currents, but you may need to find a current when you don't even know the direction in which it is flowing. Using this convention forces the correct answer!

It is vital to make sure that your signs are correct.

The voltage law states:

The algebraic sum of the EMFs and potential differences acting around any loop is equal to zero.

This law draws a very definite distinction between EMFs and potential differences. EMFs are *sources* of electrical energy (such as batteries), whereas potential differences are the *voltages dropped* across components. Another way of stating the law is to say that the algebraic sum of the EMFs must equal the algebraic sum of the potential drops around the loop. Again, you could consider this to be a conservation of voltage. See Fig. 1.3.

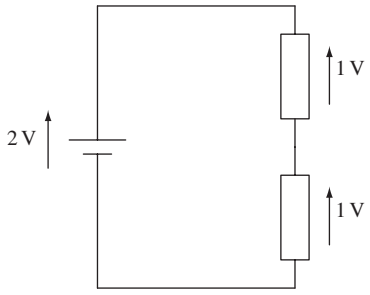


Fig. 1.3 *Summation of potentials within a loop (Kirchhoff's voltage law)*

Resistors in series and parallel

If we had a *network* of resistors, we might want to know what the total resistance was between terminals A and B. See Fig. 1.4.

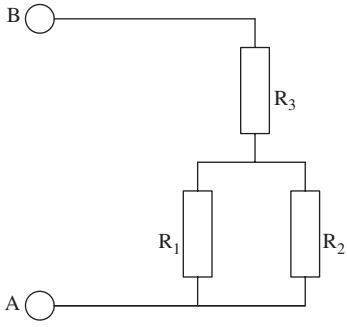


Fig. 1.4 *Series/parallel resistor network*

We have three resistors: R_1 is in *parallel* with R_2, and that *combination* is in *series* with R_3.

As with all problems, the thing to do is to break it down into its simplest parts. If we had some means of determining the value of resistors in series, we could use it to calculate the value of R_3 in series with the combination of R_1 and R_2, but as we don't yet know the value of the parallel combination, we must find this *first*. This question of order is most important, and we will return to it later.

If the two resistors (or any other component, for that matter) are in parallel, then they must have the same voltage drop across them. Ohm's law might therefore be a useful starting point.

$$I = \frac{V}{R}$$

Using Kirchhoff's current law, we can state that:

$$I_{\text{total}} = I_{\text{R1}} + I_{\text{R1}} + \cdots$$

So:

$$\frac{V}{R_{\text{parallel}}} = \frac{V}{R_1} + \frac{V}{R_2} + \cdots$$

Dividing by V:

$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2} \cdots$$

The reciprocal of the total parallel resistance is equal to the sum of the reciprocals of the individual resistors.

For the special case of only two resistors, we can derive the equation below:

$$R_{\text{parallel}} = \frac{R_1 R_2}{R_1 + R_2}$$

This is often known as 'product over sum', and whilst it is useful for mental arithmetic, it is slow to use on a calculator (more keystrokes).

Now that we have cracked the parallel problem, we need to crack the series problem.

First, we will simplify the circuit. We can now calculate the total resistance of the parallel combination and replace it with one resistor of that value – an *equivalent* resistor. See Fig. 1.5.

Using the voltage law, the sum of the potentials across the resistors must be equal to the driving EMF:

$$V_{\text{total}} = V_{\text{R1}} + V_{\text{R2}} + \cdots$$

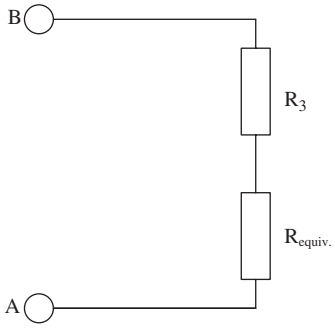


Fig. 1.5 *Simplification of Fig. 1.4 using an equivalent resistor*

Using Ohm's law:

$$V_{\text{total}} = IR_1 + IR_2 + \cdots$$

But if we are trying to create an equivalent resistor, whose value is equal to the combination, we could say:

$$IR_{\text{total}} = IR_1 + IR_2 + \cdots$$

Hence:

$$R_{\text{series}} = R_1 + R_2 + \cdots$$

The total resistance of a combination of series resistors is equal to the sum of their individual resistances.

Using the parallel and series equations, we are now able to calculate the total resistance of *any* network. See Fig. 1.6.

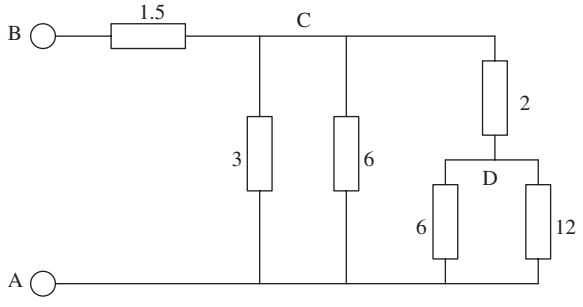


Fig. 1.6

Now this may look horrendous, but it is not a problem if we attack it logically. The hardest part of the problem is not wielding the equations or numbers, but where to start.

We want to know the resistance looking into the terminals A and B, but we don't have any rules for finding this directly, so we must look for a point where we can apply our rules. We can only apply one rule at a time, so we look for a combination of components made up *only* of series *or* parallel components.

In this example, we find that between node A and node D, there are *only* parallel components. We can calculate the value of an equivalent resistor and substitute it back into the circuit:

$$\begin{aligned} R_{\text{parallel}} &= \frac{\text{product}}{\text{sum}} \\ &= \frac{6 \times 12}{6 + 12} \\ &= 4\,\Omega \end{aligned}$$

We redraw the circuit. See Fig. 1.7.

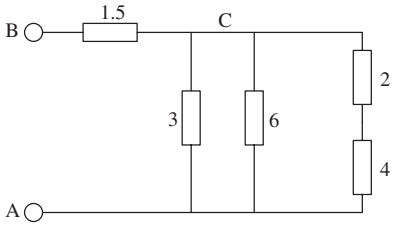


Fig. 1.7

Looking again, we see that now, the *only* combinations made up of series *or* parallel components are between node A and node C, but we have a choice – either the series combination of the 2 Ω and 4 Ω, or the parallel combination of the 3 Ω and 6 Ω. The one to go for is the series combination. This is because it will result in a single resistor that will then be in parallel with the 3 Ω and 6 Ω resistors. We can cope with the three parallel resistors later:

$$\begin{aligned} R_{\text{series}} &= R_1 + R_2 \\ &= 4 + 2 \\ &= 6\,\Omega \end{aligned}$$

We redraw the circuit. See Fig. 1.8.

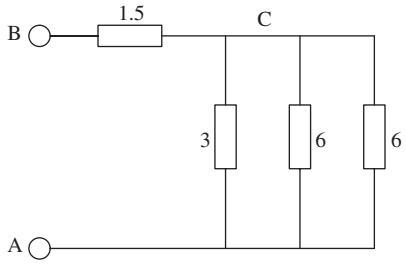


Fig. 1.8

We now see that we have three resistors in parallel:

$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

$$= \frac{1}{3} + \frac{1}{6} + \frac{1}{6}$$

$$= \frac{2}{3}$$

Hence:

$$R_{\text{parallel}} = \frac{3}{2} = 1.5\,\Omega$$

We have now reduced the circuit to two 1.5 Ω resistors in series, and so the total resistance is 3 Ω.

This took a little time, but it demonstrated some useful points that will enable you to analyse networks much faster the second time around:

- The critical stage is choosing the starting point.
- The starting point is generally as far away from the terminals as it is possible to be.
- The starting point is made up of a combination of *only* series *or* parallel components.
- Analysis tends to proceed outwards from the starting point towards the terminals.
- Redrawing the circuit helps. You may even need to redraw the original circuit if it doesn't make sense to you. Redrawing as analysis progresses reduces confusion and errors – do it!

Potential dividers

Figure 1.9 shows a potential divider. This could be made up of two discrete resistors or it could be the moving wiper of a volume control. As before, we will suppose that a current I flows through the two resistors. We want to know the ratio of the output voltage to the input voltage. See Fig. 1.9.

$$V_{out} = IR_2$$
$$V_{in} = IR_1 + IR_2$$
$$= I(R_1 + R_2)$$

Hence:

$$\frac{V_{out}}{V_{in}} = \frac{IR_2}{I(R_1 + R_2)}$$

$$= \frac{R_2}{R_1 + R_2}$$

This is a very important result and, used intelligently, can solve virtually anything.

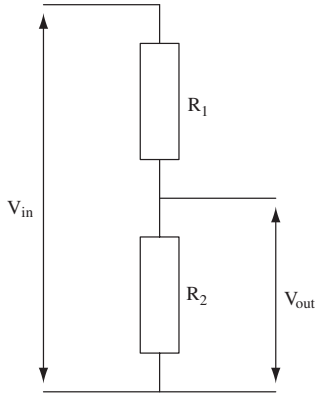


Fig. 1.9 *Potential divider*

Equivalent circuits

We have looked at networks of resistors and calculated equivalent *resistances*. Now we will extend the idea to equivalent *circuits*. This is a tremendously powerful concept for circuit analysis.

It should be noted that this is not the only method, but it is usually the quickest, and kills 99% of all known problems. Other methods include Kirchhoff's laws combined with lots of simultaneous equations and the superposition theorem. These methods may be found in standard texts, but they tend to be cumbersome, so we will not discuss them here.

The Thévenin equivalent circuit

When we looked at the potential divider, we were able to calculate the ratio of output voltage to input voltage. If we were now to connect a battery across the input terminals, we could calculate the output voltage. Using our earlier tools, we could also calculate the total resistance looking into the output terminals. As before, we could then redraw the circuit, and the result is known as the *Thévenin equivalent circuit*. If two black boxes were made, one containing the original circuit, and the other the Thévenin equivalent, you would not be able to tell from the output terminals which was which. The concept is simple to use and can break down complex networks quickly and efficiently. See Fig. 1.10.

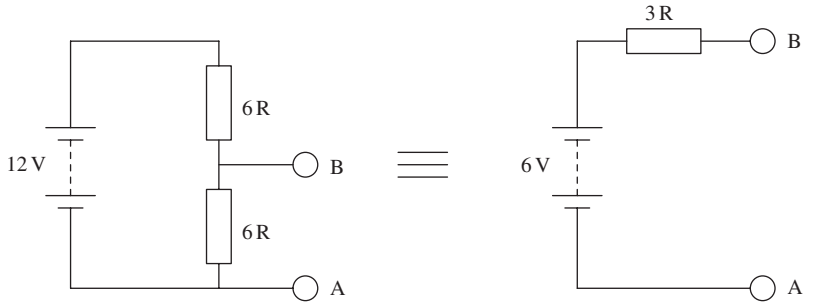


Fig. 1.10 *A 'black box' network, and its Thévenin equivalent circuit*

This is a simple example to demonstrate the concept. First, we find the equivalent resistance, often known as the *output resistance*. Now, in the world of equivalent circuits, batteries are perfect voltage sources, they have zero *internal resistance* and look like a short circuit when we consider their

resistance. Therefore, we can ignore the battery, or replace it with a piece of wire whilst we calculate the resistance of the total circuit:

$$R_{\text{output}} = \frac{R_1 R_2}{R_1 + R_2}$$
$$= 3\,\Omega$$

Next, we need to find the output voltage. We will use the potential divider equation:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{R_2}{R_1 + R_2}$$
$$= \frac{1}{2}$$

So:

$$V_{\text{out}} = \frac{1}{2} \cdot V_{\text{in}}$$
$$= 6\,\text{V}$$

Now for a much more complex example: this will use all the previous techniques and really demonstrate the power of Thévenin. See Fig. 1.11.

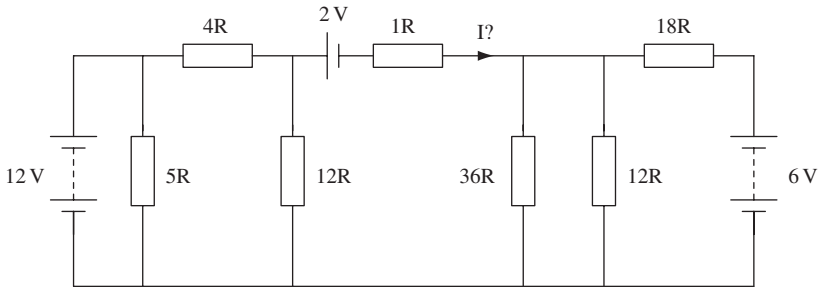


Fig. 1.11

For some obscure reason, we want to know the current flowing in the 1 Ω resistor. The first thing to do is to redraw the circuit. Before we do this we can observe that the 5 Ω resistor in parallel with the 12 V battery is irrelevant. Yes, it will draw current from the battery, but it doesn't affect the operation of

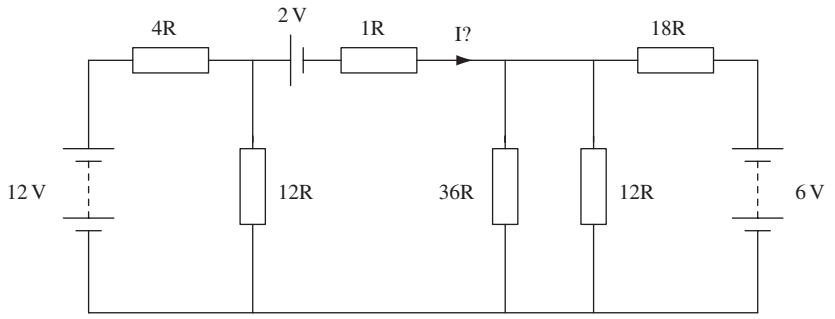


Fig. 1.12

the rest of the circuit. A short circuit in parallel with 5 Ω is *still* a short circuit, so we will throw it away. See Fig. 1.12.

Despite our best efforts, this is still a complex circuit, so we need to break it down into modules that we recognize. Looking at the left-hand side, we see a battery with a 4 Ω and 12 Ω resistor which looks suspiciously like the simple problem that we saw earlier, so let's break the circuit there and make an equivalent circuit. See Fig. 1.13.

Using the potential divider rule:

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} \cdot V_{\text{in}}$$

$$= 9\,\text{V}$$

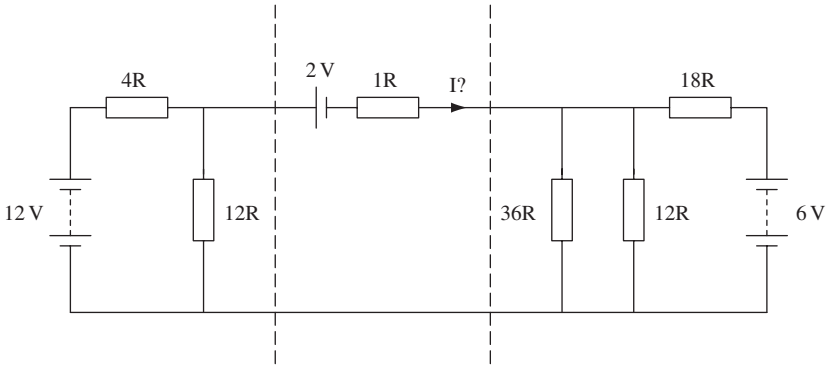


Fig. 1.13

Using 'product over sum':

$$R_{out} = \frac{R_1 R_2}{R_1 + R_2}$$
$$= 3\,\Omega$$

Looking at the right-hand side, we can perform a similar operation to the right of the dotted line.

First, we find the resistance of the parallel combination of the 36 Ω and 12 Ω resistors, which is 9 Ω. We now have a potential divider, whose output resistance is 6 Ω, and the Thévenin voltage is 2 V.

Now we redraw the circuit. See Fig. 1.14.

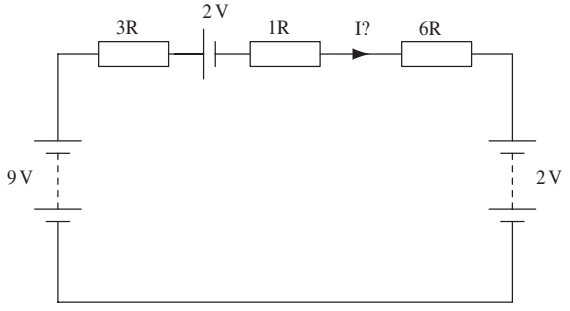


Fig. 1.14

We can make a few observations at this point. First, we have three batteries in series, why not combine them into one battery? There is no reason why we should not do this provided that we take note of their polarities. Similarly, we can combine some or all of the resistors. See Fig. 1.15.

The problem now is trivial, and a simple application of Ohm's law will solve it. We have a total resistance of 10 Ω, and a 5 V battery, so the current must be ½ A.

Useful points to note:

- Look for components that are irrelevant, such as resistors directly across battery terminals.
- Look for potential dividers on the outputs of batteries and 'Thévenize' them. Keep on doing so until you meet the next battery.
- Work from battery terminals outwards.
- Keep calm, and try to work neatly – it will save mistakes later.

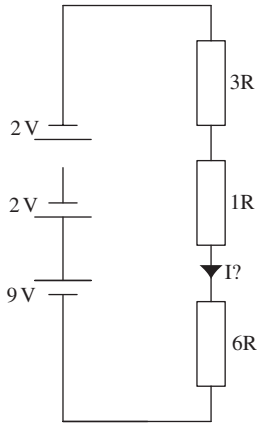


Fig. 1.15

Although it is possible to solve most problems using a Thévenin equivalent circuit, sometimes a Norton equivalent is more convenient.

The Norton equivalent circuit

The Thévenin equivalent circuit was a perfect voltage source in *series* with a resistance, whereas the *Norton* equivalent circuit is a perfect *current* source in *parallel* with a resistance. See Fig. 1.16.

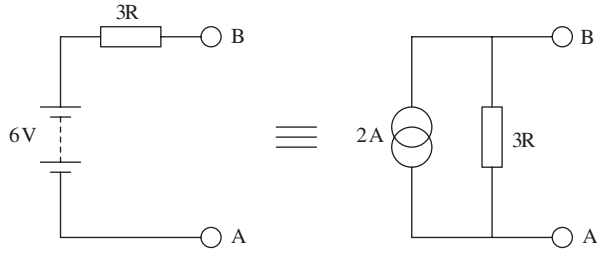


Fig. 1.16 *The Norton equivalent circuit*

We can easily convert from a Norton source to a Thévenin source, or vice versa, because the resistor has the same value in both cases. We find the value of the current source by short circuiting the output of the Thévenin source and calculating the resulting current – this is the Norton current.

To convert from a Norton source to a Thévenin source, we leave the source open circuit and calculate the voltage developed across the Norton resistor – this is the Thévenin voltage.

For the vast majority of problems, the Thévenin equivalent will be quicker, mostly because we become used to thinking in terms of voltages that can easily be measured by a meter, or viewed on an oscilloscope. Occasionally, a problem will arise that is intractable using Thévenin, and converting to a Norton equivalent causes the problem to solve itself. Norton problems usually involve the summation of a number of currents, when the only other solution would be to resort to Kirchhoff and simultaneous equations.

Units and multipliers

All the calculations up to this point have been arranged to use convenient values of voltage, current and resistance. In the real world, we will not be so fortunate, and to avoid having to use scientific notation, which takes longer to write, and is virtually unpronounceable, we will prefix our units with multipliers.[1]

Prefix	Abbreviation	Multiplies by
yocto	y	10^{-24}
zepto	z	10^{-21}
atto	a	10^{-18}
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}
kilo	k	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}
exa	E	10^{18}
zetta	Z	10^{21}
yotta	Y	10^{24}

Note that the *case* of the prefix is important; there is a large difference between 1 mΩ and 1 MΩ. Electronics uses a very wide range of values; small signal pentodes have anode to grid capacitances measured in fF (F = farad, the unit of capacitance), and petabyte data stores are already common. Despite this, for day-to-day electronics use, we only need to use pico to mega.

Electronics engineers commonly abbreviate further, and you will often hear a 22 pF (picofarad) capacitor referred to as 22 'puff', whilst the 'ohm' is commonly dropped for resistors, and 470 kΩ (kilo-ohm) would be pronounced as 'four-seventy-kay'.

A rather more awkward abbreviation that arose before high resolution printers became available (early printers and typewriters couldn't print 'μ'), is the abbreviation of μ (micro) to m. This abbreviation persists, particularly in American text, and you will occasionally see a 10 mF capacitor specified, although the context makes it clear that what is actually meant is 10 μF. For this reason, true 10 mF capacitors are invariably specified as 10 000 μF.

Unless an equation states otherwise, you may assume that it uses the base physical units, so an equation involving capacitance and time constants would expect you to express capacitance in farads and time in seconds. Thus, $75\,\mu s = 75 \times 10^{-6}$ seconds, and the value of capacitance determined by an equation might be $2.2 \times 10^{-10}\,F = 220\,pF$.

The decibel

The human ear spans a vast dynamic range from the near silence heard in an empty recording studio to the deafening noise of a nearby pneumatic drill. If we were to plot this range linearly on a graph, the quieter sounds would hardly be seen, whereas the difference between the noise of the drill and a jet engine would be given a disproportionate amount of room on the graph. What we need is a graph that gives an equal weighting to *relative* changes in level of both quiet and loud sounds. By definition, this implies a logarithmic scale on the graph, but electronics engineers went one better and invented a logarithmic ratio known as the *decibel* (dB) which was promptly hijacked by the acoustical engineers. (The fundamental unit is the Bel, but this is inconveniently large, so the decibel is more commonly used, and the capital b is usually dropped.)

The dB is *not* an absolute quantity. It is a *ratio*, and it has one formula for use with currents and voltages, and another for powers:

$$\text{dB} = 20 \log_{10}\left(\frac{V_1}{V_2}\right) = 20 \log_{10}\left(\frac{I_1}{I_2}\right) = 10 \log_{10}\left(\frac{P_1}{P_2}\right)$$

The reason for this is that $P \propto V^2$ or I^2, and with logarithms, multiplying the logarithm by 2 is the same as squaring the original number. Using a different formula to calculate dBs when using powers ensures that the resulting dBs are equivalent, irrespective of whether they were derived from powers or voltages.

This might seem complicated when all we wanted to do was to describe the difference in two signal levels, but the dB is a very handy unit.

Useful common dB values are:

dB	V_1/V_2	P_1/P_2
0	1	1
3	$\sqrt{2}$	2
6	2	4
20	10	100

Because dBs are derived from logarithms, they obey all the rules of logarithms, and adding dBs is the same as multiplying the ratios that generated them. Note that dBs can be negative, implying loss, or a drop in level.

For example, if we had two cascaded amplifiers, one with a voltage gain of ½ and the other with a voltage gain of 10, then by *multiplying* the individual gains, the combined voltage gain would be 5. Alternatively, we could find the gain in a dB by saying that one amplifier had −6 dB of gain whilst the other had 20 dB, and *adding* the gains in a dB to give a total gain of 14 dB.

When *designing* amplifiers, we will not often use the above example, as absolute voltages are often more convenient, but we frequently need dBs to describe filter and equalization curves.

Alternating current

All the previous techniques have used *direct current* (DC), where the current is constant and flows in one direction only. Listening to DC is not very interesting, so we now need to look at *alternating currents* (AC).

All of the previous techniques of circuit analysis can be applied equally well to AC signals.

The sine wave

The *sine* wave is the simplest possible alternating signal, and its equation is:

$$v = V_{peak} \sin(\omega t + \theta)$$

where: v = the instantaneous value at time t
V_{peak} = the peak value
ω = angular frequency in radians/second ($\omega = 2\pi f$)
t = the time in seconds
θ = a constant phase angle

These mathematical concepts are shown on the diagram. See Fig. 1.17.

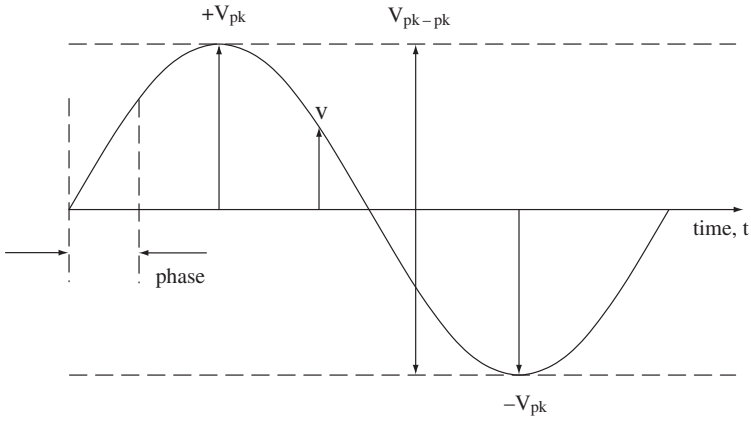


Fig. 1.17

There is a convention in use in this equation. *Upper* case letters denote DC, or constant values, whereas *lower* case letters denote the instantaneous AC, or changing, value. It is a form of shorthand to avoid having to specify separately that a quantity is AC or DC. It would be nice to say that this convention is rigidly applied, but it is often neglected, and the context of the symbols usually makes it clear whether the quantity is AC or DC.

In electronics, the word 'peak' (pk) has a very precise meaning and, when used to describe an AC waveform, it means the voltage from zero volts to the

peak voltage reached, either positive or negative. Peak-to-peak (pk–pk) means the voltage from positive peak to negative peak and, for a symmetrical waveform, $V_{pk-pk} = 2V_{pk}$.

Although electronics engineers habitually use ω to describe frequency, they do so only because calculus requires that they work in radians. Since $\omega = 2\pi f$, we can rewrite the equation as:

$$v = V_{peak} \sin(2\pi ft + \theta)$$

If we now inspect this equation, we see that apart from time t, we could vary other constants before we allow time to change and determine the waveform. We can change V_{peak}, and this will change the *amplitude* of the sine wave, or we can change f, and this will change the *frequency*. The inverse of frequency is *period*, which is the time taken for one full cycle of the waveform to occur:

$$\text{period}(T) = \frac{1}{f}$$

If we listen to a sound that is a sine wave, and change the amplitude, this will make the sound louder or softer, whereas varying frequency changes the pitch. If we vary θ (phase), it will sound the same if we are listening to it, and unless we have an external reference, the sine wave will look exactly the same viewed on an oscilloscope. Phase becomes significant if we compare one sine wave with another sine wave *of the same frequency*, or a harmonic of that frequency. Attempting to compare phase between waveforms of unrelated frequencies is meaningless.

Now that we have described sine waves, we can look at them as they would appear on the screen of an oscilloscope. See Fig. 1.18.

Sine waves A and B are identical in amplitude, frequency and phase. Sine wave C has lower amplitude, but frequency and phase are the same. Sine wave D has the same amplitude, but double the frequency. Sine wave E has identical amplitude and frequency to A and B, but the phase θ has been changed.

Sine wave F has been *inverted*. Although, for a sine wave, we cannot see the difference between a 180° phase change and an inversion, for asymmetric waveforms, there is a distinct difference. We should therefore be very careful if we say that two waveforms are 180° out of phase with each other that we do not actually mean that one is inverted with respect to the other.

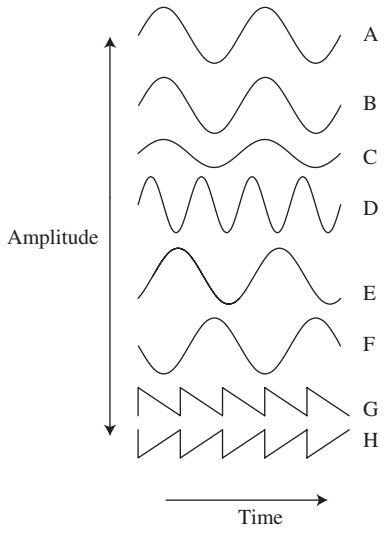


Fig. 1.18

The *sawtooth* waveforms G and H are inverted, and it can be seen that this is completely different from a 180° phase change. (Strictly, the term 'phase splitter' is incorrect for this reason, but the more technically correct description 'phase invertor' is less common in the UK than the USA.)

The transformer

When the electric light was introduced as an alternative to the gas mantle, there was a great debate as to whether the distribution system should be AC or DC. The outcome was settled by the enormous advantage of the *transformer* which could step up, or step down, the voltage of an AC supply. The DC supply could not be manipulated in this way, and evolution took its course.

A transformer is essentially a pair of electrically insulated windings that are magnetically coupled to each other, usually on an iron core. They vary from fingernail size to the size of a large house, depending on power rating and operating frequency, with high frequency transformers being smaller. The symbol for a transformer is modified depending on the core material. Solid lines indicate a laminated iron core, dotted lines denote a dust core, whilst an air core has no lines. See Fig. 1.19.

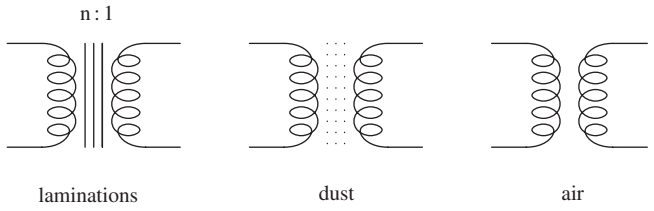


Fig. 1.19 *Transformer symbols*

The perfect transformer changes one AC voltage to another, more convenient, voltage with no losses whatsoever; all of the power at the input is transferred to the output:

$$P_{\text{in}} = P_{\text{out}}$$

Having made this statement, we can now derive some useful equations:

$$V_{\text{in}} \cdot I_{\text{in}} = V_{\text{out}} \cdot I_{\text{out}}$$

Rearranging:

$$\frac{V_{\text{in}}}{V_{\text{out}}} = \frac{I_{\text{out}}}{I_{\text{in}}} = n$$

The new constant n is very important, and is the ratio between the number of turns on the input winding and the number of turns on the output winding of the transformer. Habitually, when we talk about transformers, the input winding is known as the *primary* and the output winding is the *secondary*:

$$n(\text{turns ratio}) = \frac{\text{number of primary turns}}{\text{number of secondary turns}}$$

Occasionally, a transformer may have a winding known as a *tertiary* winding, which usually refers to a winding used for feedback or monitoring, but it is more usual to refer to multiple primaries and secondaries.

When the perfect transformer steps voltage down, perhaps from 240 V to 12 V, the current ratio is stepped up, and for each ampere of primary current, 20 A must be drawn by the secondary. This implies that the resistance of the

load on the secondary is different from that seen looking into the primary. If we substitute Ohm's law into the conservation of power equation:

$$\frac{V_{\text{primary}}^2}{R_{\text{primary}}} = \frac{V_{\text{secondary}}^2}{R_{\text{secondary}}}$$

$$\frac{R_{\text{secondary}}}{R_{\text{primary}}} = \left(\frac{V_{\text{primary}}}{V_{\text{secondary}}}\right)^2 = n^2$$

The transformer changes resistances by the *square* of the turns ratio. This will become very significant when we use audio transformers and need to match the resistance of a loudspeaker to the output valves.

As an example, an output transformer with a primary to secondary turns ratio of 31.6:1 would allow the output valves to see the 8 Ω loudspeaker as an 8 kΩ load, whereas the loudspeaker sees the Thévenin output resistance of the output valves stepped down by an identical amount.

The concept of looking into a device in one direction, and seeing one thing, whilst looking in the opposite direction, and seeing another, is very powerful, and we will use it frequently when we investigate simple amplifier stages.

Practical transformers are not perfect, and we will investigate their imperfections in greater detail in Chapter 4.

Capacitors, inductors and reactance

Previously, when we analysed circuits, they were composed purely of resistors and voltage or current sources.

We now need to introduce two new components: *capacitors* and *inductors*. Capacitors have the symbol C, and the unit of capacitance is the *farad* (F). 1 F is an extremely large capacitance, and more common values range from a few pF to tens of thousands of μF. Inductors have the symbol *L*, and the unit of inductance is the *henry* (H). The henry is quite a large unit, and common values range from a few μH to tens of H. Although the henry, and particularly the farad, are rather large for our very specialized use, their size derives from the fundamental requirement for a coherent system of units. A coherent system allows units (such as the farad) to be derived from base units (such as the ampere) without needing any scaling factors.

The simplest capacitor is made of a pair of separated plates, whereas an inductor is a coil of wire, and this physical construction is reflected in their graphic symbols. See Fig. 1.20.

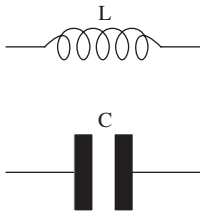


Fig. 1.20 *Inductor and capacitor symbols*

Resistors had resistance, whereas capacitors and inductors have *reactance*. Reactance is the AC equivalent of resistance – it is still measured in ohms, and is given the symbol X. We will often have circuits where there is a combination of inductors and capacitors, so it is normal to add a subscript to denote which reactance is which:

$$X_C = \frac{1}{2\pi fC}$$
$$X_L = 2\pi fL$$

Looking at these equations, we see that reactance changes with frequency and with the value of the component. We can plot this relationship as a graph. See Fig. 1.21.

An inductor has a reactance of zero at zero frequency. More intuitively, it is a short circuit at DC. As we increase frequency, its reactance rises.

A capacitor has infinite reactance at zero frequency. It is open circuit at DC. As frequency rises, reactance falls.

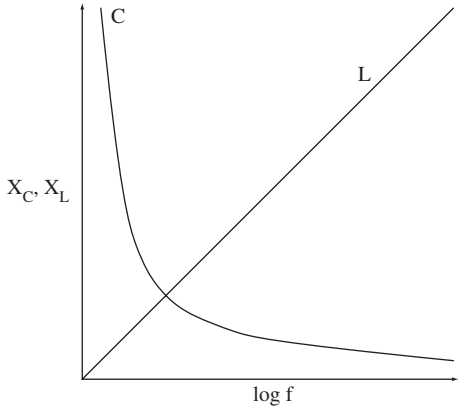


Fig. 1.21 *Reactance of inductor and capacitor against frequency*

A circuit made of only one capacitor, or one inductor, is not very interesting, and we might want to describe the behaviour of a circuit made up of resistance and reactance, such as a moving coil loudspeaker. See Fig. 1.22.

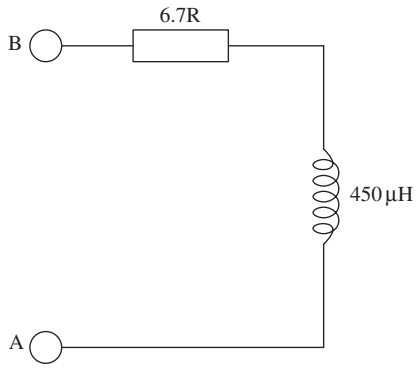


Fig. 1.22

We have a combination of resistance and inductive reactance, but at the terminals A, B we see neither a pure reactance nor a pure resistance, but a combination of the two factors known as *impedance*.

In a traditional electronics book we would now lurch into the world of vectors, phasors and complex number algebra. Whilst fundamental AC theory is essential for electronics engineers who have to pass examinations, we cannot justify the mental trauma needed to cover the topic in depth, so we will simply pick out useful results that are relevant to our highly specialized field of interest.

Filters

We mentioned that reactance varies with frequency. This property can be used to make a *filter* that allows some frequencies to pass unchecked, whilst others are attenuated. See Fig. 1.23.

All filters are based on potential dividers. In this filter, the upper leg of the potential divider is a capacitor, whereas the lower leg is a resistor. We stated earlier that a capacitor is an open circuit at DC. This filter will therefore have infinite attenuation at DC – it *blocks* DC. At infinite frequency, the capacitor is a short circuit, and the filter passes the signal with no attenuation, so the filter is known as a *high-pass* filter. See Fig. 1.24.

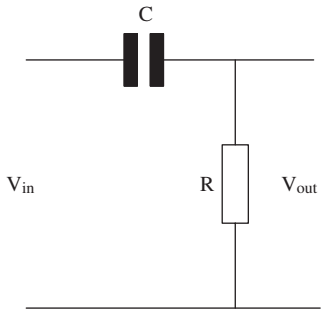


Fig. 1.23 *CR high-pass filter*

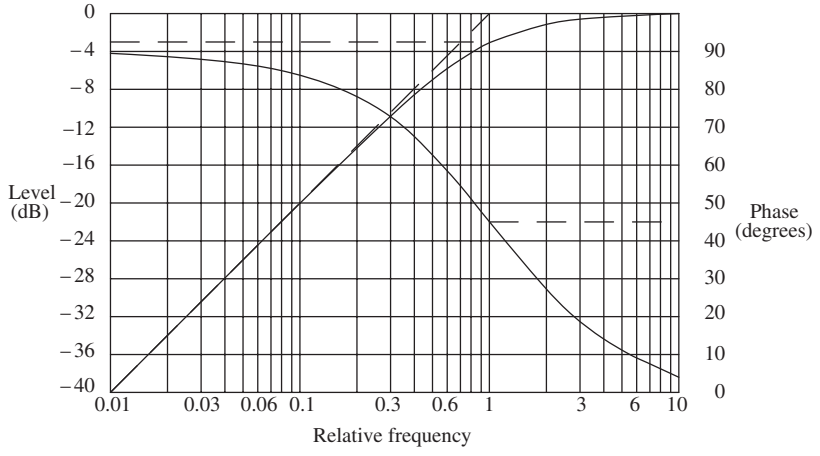


Fig. 1.24 *Frequency response of the CR high-pass filter*

Frequency is plotted on a logarithmic scale to encompass the wide range of values without cramping. When we needed a logarithmic unit for amplitude ratios, we invented the dB, but a logarithmic unit for frequency already existed, so engineers stole the *octave* from the musicians. An octave is simply a halving or doubling of frequency, and corresponds to eight 'white keys' on a piano keyboard.

The curve has three distinct regions: the *stop-band, cut-off* and the *pass-band*.

The stop-band is the region where signals are stopped, or attenuated. In this filter, the attenuation is inversely proportional to frequency, and we can see that at a sufficiently low frequency, the shape of the curve in this region becomes a straight line. If we were to measure the slope of this line, we would find that it tends towards 6 dB/octave.

Note that the *phase* of the output signal changes with frequency, with a maximum change of 90° when the curve finally reaches 6 dB/octave.

This slope is very significant, and all filters with only one reactive element have an ultimate slope of 6 dB/octave. As we add more reactive elements, we can achieve a higher slope, so filters are often referred to by their *order*, which is the number of reactive elements contributing to the slope. A third order filter would have three reactive elements, and its ultimate slope would therefore be 18 dB/octave.

Although the curve reaches an ultimate slope, the behaviour at cut-off is of interest, not least because it allows us to say at what frequency the filter begins to take effect. On the diagram, a line was drawn to determine the ultimate slope. If this is extended until it intersects with a similar line drawn continued from the pass-band attenuation, the point of intersection is the filter cut-off frequency. (You will occasionally see idealized filter responses drawn in this way, but this does not imply that the filter response actually changes abruptly from pass-band to stop-band.)

If we now drop a line down to the frequency axis from the cut-off point, it passes through the curve, and the filter response at this point is 3 dB down on the pass-band value. The cut-off frequency is therefore also known as $f_{-3\,\mathrm{dB}}$, or the -3 dB point, and at this point the phase curve is at its steepest, with a phase change of 45°.

Second order filters, and above, have considerable freedom in the way that the transition from pass-band to stop-band is made, and so the *class* of filter is often mentioned in conjunction with names like Bessel, Butterworth and Chebychev, in honour of their originators.

Although we initially investigated a high-pass *CR* filter, other combinations can be made using one reactive component and a resistor. See Fig. 1.25.

We now have a pair of high-pass filters, and a pair of *low-pass* filters; the low-pass filters have the same slope, and cut-off frequency can be found from the graph in the same way.

Now that we are familiar with the shape of the curves of these simple filters, referring to them by their cut-off frequency and slope is more convenient (the word 'ultimate' is commonly neglected). For these simple filters, the equation for cut-off frequency is the same whether the filter is high pass or low pass. For a *CR* filter:

$$f_{-3\,\mathrm{dB}} = \frac{1}{2\pi CR}$$

And for an *LR* filter:

$$f_{-3\,\mathrm{dB}} = \frac{R}{2\pi L}$$

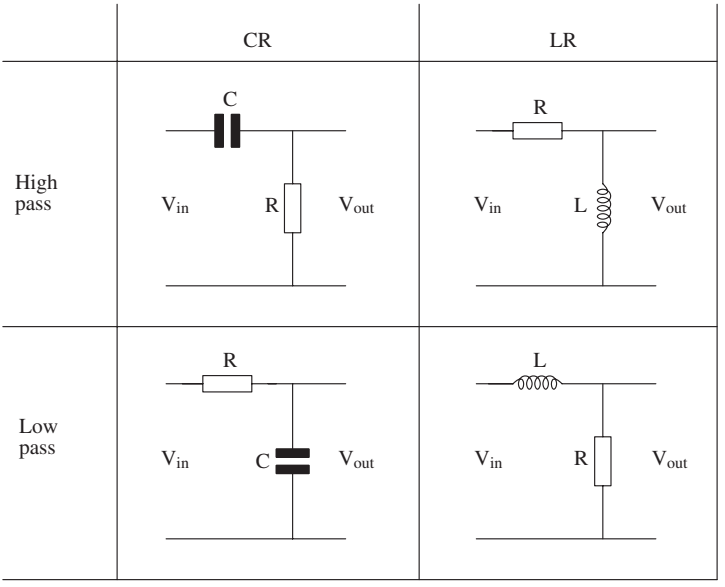


Fig. 1.25

Time constants

In audio, simple filters or *equalization* networks are often described in terms of their *time constants*. These have a very specialized meaning that we will touch upon later, but in this context, they are simply used as a shorthand form of describing a first order filter that allows component values to be calculated quickly.

For a *CR* network, the time constant τ (tau) is:

$$\tau = CR$$

For an *LR* network:

$$\tau = \frac{L}{R}$$

Because it is a *time* constant, the units of τ are seconds, but audio time constants are habitually given in μs. We can easily calculate the cut-off frequency of the filter from its time constant τ:

$$f = \frac{1}{2\pi\tau}$$

Note that τ is quite distinct from period, which is given the symbol T.

Examples of audio time constants: RIAA record equalization for vinyl records – 3180 μs, 318 μs, 75 μs; analogue FM broadcast radio de-emphasis – UK 50 μs, USA 75 μs.

A 75 μs HF de-emphasis circuit needs a low-pass filter, usually *CR*, so we choose a pair of component values whose product equals 75 μs, 1 nF and 75 kΩ would do nicely. See Fig. 1.26.

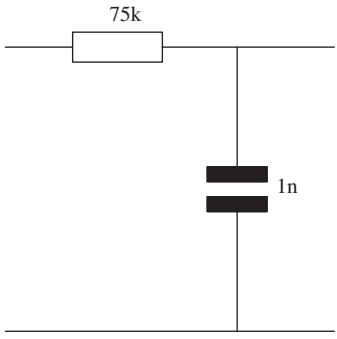


Fig. 1.26 *A 120 μs de-emphasis network*

Resonance

So far, we have made filters using only one reactive component, but if we make a network using a capacitor *and* an inductor, we find that we have a *resonant* circuit. Resonance occurs everywhere in the natural world, from the sound of a tuning fork, to the bucking and twisting of the Tacoma Narrows bridge. (A bridge that finally collapsed during a storm on 7 November 1940 because the wind excited a structural resonance.) A resonant electronic circuit is shown in Fig. 1.27.

If we were to sweep the frequency of the source, whilst measuring the current drawn, we would find that at the resonant frequency, the current would rise to a maximum determined purely by the resistance of the resistor. The circuit would appear as if the other components were not there. We could then plot a graph of current against frequency. See Fig. 1.28.

The sharpness and height of this peak is determined by the *Q* or *magnification factor* of the circuit:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}}$$

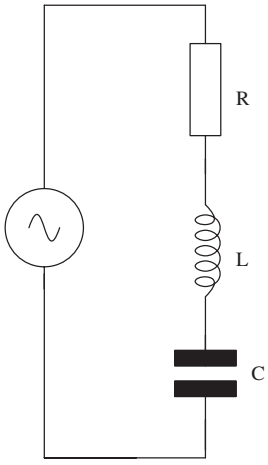


Fig. 1.27 *Series resonant circuit*

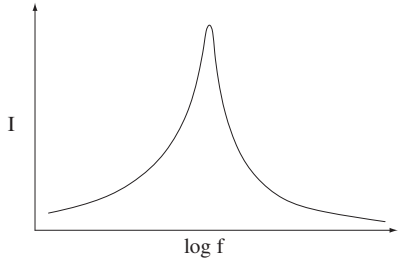


Fig. 1.28 *Current against frequency for series resonant circuit*

This shows us that a small resistance can cause a high Q, and this will be very significant later. The frequency of resonance is:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

Our first resonant circuit was a *series* resonant circuit, but *parallel* resonance, where the *total* current falls to a minimum at resonance, is also possible. See Fig. 1.29.

If $Q > 5$, the above equations are reasonably accurate for parallel resonance. We will not often worry greatly about the accuracy of resonant

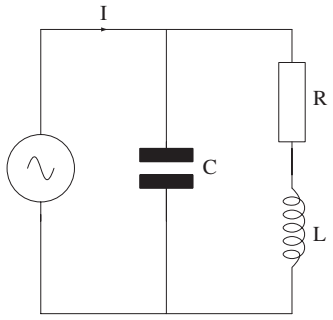


Fig. 1.29 *Parallel resonant circuit*

calculations, since we do not want resonances in audio, and we will do our best to remove them.

RMS and power

We mentioned power earlier, when we investigated the flow of current through a lightbulb using a 240 V battery. Mains electricity is AC and has recently been respecified in the UK to be $230\,V_{AC} + 10\% - 6\%$ at $50\,Hz \pm 1\,Hz$, but how do we define the 230 V?

If we had a valve heater filament, it would be most useful if it could operate equally well from AC or DC. As far as the valve is concerned, AC electricity will heat the filament equally well, just so long as we apply the correct voltage:

> *The RMS voltage of any waveform is equivalent to the DC voltage having the same heating effect as the original waveform.*

RMS is short for *R*oot of the *M*ean of the *S*quares, which refers to the method of calculating the value. Fortunately, the ratios of V_{RMS} to V_{peak} have been calculated for the common waveforms, and in audio design, we are mostly concerned with the sine wave, for which:

$$V_{peak} = \sqrt{2} \cdot V_{RMS}$$

All sinusoidal AC voltages are given in V_{RMS} unless specified otherwise, so a heater designed to operate at $6.3\,V_{AC}$ would work equally well connected to $6.3\,V_{DC}$.

We have only mentioned RMS voltages, but we can equally well have RMS currents, in which case:

$$P = V_{\text{RMS}} \cdot I_{\text{RMS}}$$

There is no such thing as an RMS watt!
Please refer to the definition of RMS.

The square wave

Until now, all of our dealings have been with sine waves, which are pure tones. When we listen to music, we do not hear pure tones, instead we hear a fundamental with various proportions of harmonics whose frequencies are arithmetically related to the frequency of the fundamental. We are able to distinguish between one instrument and another because of the differing proportions of the harmonics, and because of the transient at the beginning of each note.

A useful waveform for testing amplifiers quickly would have many harmonics and a transient component. The square wave has precisely these properties, because it is composed of a fundamental frequency plus odd harmonics whose amplitudes steadily decrease with frequency. See Fig. 1.30.

A square wave is thus an infinite series of harmonics, all of which must be summed from the fundamental to the infinite harmonic. We can express this

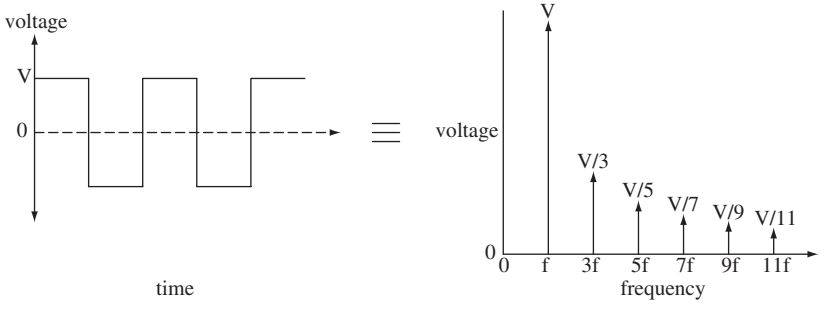


Fig. 1.30 *Square wave viewed in time and frequency*

argument mathematically as a *Fourier* series, where f is the fundamental frequency:

$$\text{square wave} = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(2n-1)\cdot 2\pi ft]}{2n-1}$$

This is a shorthand formula, but understanding the distribution of harmonics is much easier if we express them in the following form:

$$\text{square wave} \propto 1(f) + \frac{1}{3}(3f) + \frac{1}{5}(5f) + \frac{1}{7}(7f) + \frac{1}{9}(9f) + \cdots$$

We can now see that the harmonics die away very gradually, and that a 1 kHz square wave has significant harmonics well beyond 20 kHz. What is not explicitly stated by these formulae is that the relative phase of these components is critical. The square wave thus not only tests amplitude response, but also phase response.

Square waves and transients

We briefly mentioned earlier that the square wave contained a transient component. One way of viewing a square wave is to treat it as a DC level whose polarity is inverted at regular intervals. At the instant of inversion, the voltage has to change instantaneously from its negative level to its positive level, or vice versa. The abrupt change at the leading edge of the square wave is the transient, and because it occurs so quickly, it must contain a high proportion of high frequency components. Although we already knew that the square wave contained these high frequencies, it is only at the leading edge that they all sum constructively, so any change in high frequency response is seen at this leading edge.

We have considered the square wave in terms of frequency, now we will consider it as a series of transients in *time*, and investigate its effect on the behaviour of *CR* and *LR* networks.

The best way of understanding this topic is with a mixture of intuitive reasoning coupled to a few graphs. Equations *are* available, but we very rarely need to use them.

In electronics, a *step* is an *instantaneous* change in a quantity such as current or voltage, so it is a very useful theoretical concept for exploring the response of circuits to transients. We will start by looking at the voltage across a capacitor when a voltage step is applied via a series resistor. See Fig. 1.31.

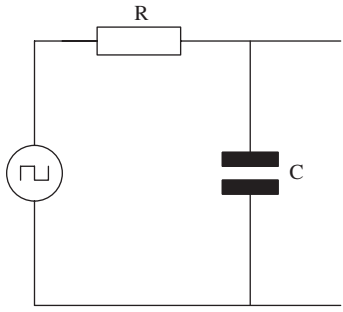


Fig. 1.31

The capacitor is initially discharged ($V_c = 0$). The step is applied and switches from 0 V to +V, an instantaneous change of voltage composed mostly of high frequencies. The capacitor has a reactance that is inversely proportional to frequency, and therefore appears as a short circuit to these high frequencies. If it is a short circuit, we cannot develop a voltage across it. The resistor therefore has the full applied voltage across it, and passes a current determined by Ohm's law. This current then flows through the capacitor, and starts charging the capacitor. As the capacitor charges, its voltage rises, until eventually, it is fully charged, and no more current flows. If no more current flows into the capacitor ($I_c = 0$), then $I_R = 0$, and so $V_R = 0$. We can plot this argument as a pair of graphs showing capacitor and resistor voltage. See Fig. 1.32.

The first point to note about these two graphs is that the shape of the curve is an *exponential* (this term will be explained further later). The second point is that when we apply a step to a *CR* circuit (and even an *LR* circuit), the current or voltage curve will *always* be one of these curves. Knowing that the curve can only be one of these two possibilities, all we need to be able to do is to choose the appropriate curve.

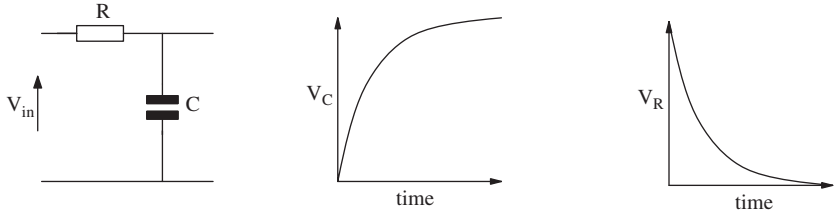


Fig. 1.32 *Exponential response of CR circuit to voltage step*

The transient edge can be considered to be of infinitely high frequency, and the capacitor is therefore a short circuit. Developing a voltage across a short circuit requires infinite current:

Infinite current is required to change the voltage across a capacitor instantaneously.

An inductor is the *dual* or inverse of a capacitor, and so an inductor has a similar rule:

An instantaneous change of current through an inductor creates an infinite voltage.

We can now draw graphs for each of the four combinations of *CR* and *LR* circuits when the same step in voltage is applied. See Fig. 1.33.

Having stated that the shape of the curves in each of the four cases is identical, we can now examine the fundamental curves in a little more detail.

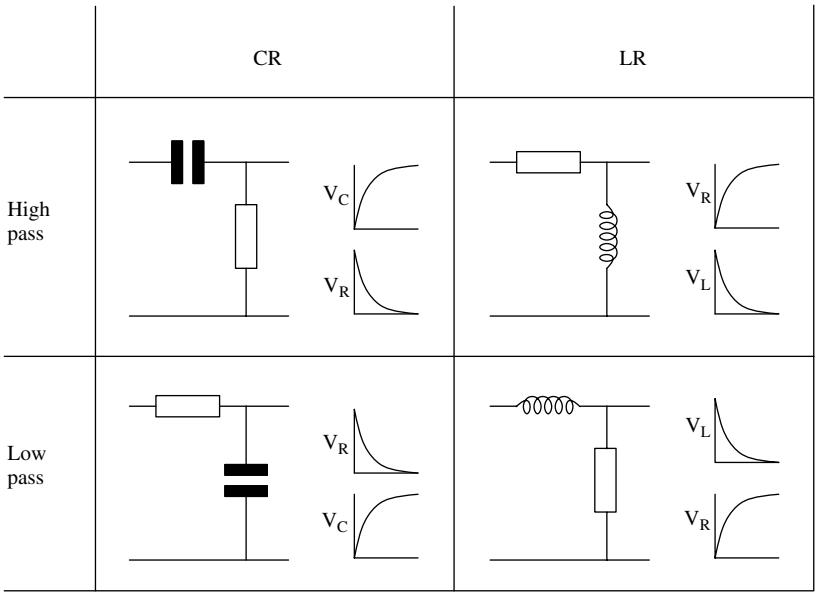


Fig. 1.33

Using the original CR circuit as an example, the capacitor will eventually charge to the input voltage, so we can draw a dotted line to represent this voltage. The voltage across the capacitor has an initial slope, and if we continue this slope with another dotted line, we will find that it intersects the first line at a time that corresponds to CR, which is the *time constant* that we met earlier. The CR time constant is defined as the time taken for the capacitor voltage to reach its final value had the initial rate of charge been maintained. See Fig. 1.34.

The equation for the falling curve is:

$$v = V \cdot \exp^{-t/\tau}$$

The equation for the rising curve is:

$$v = V\left(1 - \exp^{-t/\tau}\right)$$

Where 'exp' is the base of natural logarithms, and is the key marked 'e^x' or 'exp' on your scientific calculator. These curves derive their name because they are based on an exponential function.

We could now find what voltage the capacitor actually achieved at various times. Using the equation for the rising curve:

1τ	63%
3τ	95%
3τ	99%

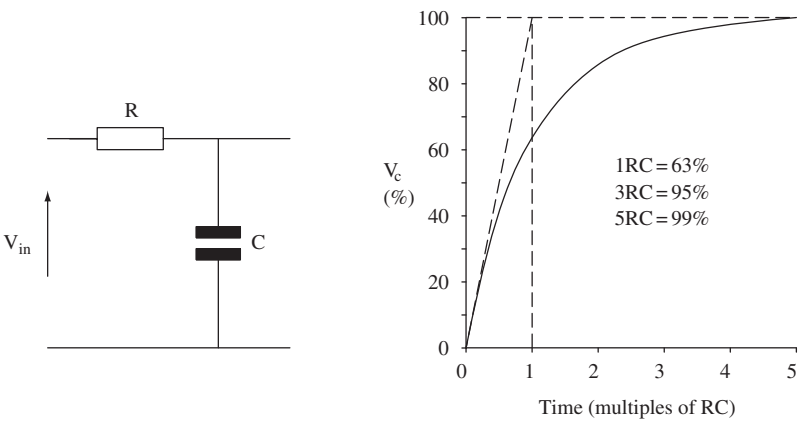


Fig. 1.34 *T and its significance to the exponential curve*

Because the curves are all the same shape, these ratios apply to all four of the *CR*, *LR* combinations. The best way to use the ratios is to first decide which way the curve is heading; the ratios then determine how much of the *change* will be achieved, and in what time.

Note that after 5τ, the circuit has very nearly reached its final position, or steady state. This is a useful point to remember when considering what will happen at switch-on to high voltage semiconductor circuits.

When we considered the response to a single step, the circuit eventually achieved a steady state because there was sufficient time for the capacitor to charge, or for the inductor to change its magnetic field. With a square wave, this may no longer be true. As was mentioned before, the square wave is an excellent waveform for testing audio amplifiers, not least because oscillators that can generate both sine and square waves are fairly cheaply available.

If we apply a square wave to an amplifier, we are effectively testing a *CR* circuit made up of a series resistance and a capacitance to ground (often known as a shunt capacitance). We should therefore expect to see some rounding of the leading edges, because some of the high frequencies are being attenuated.

If the amplifier is only marginally stable (because it contains an unwanted resonant circuit) the high frequencies at the leading edges of the square wave will excite the resonance, and we may see a damped train of oscillations following each transition.

We can also test low frequency response with a square wave. If the coupling capacitors between stages are small enough to change their charge noticeably within one half cycle of the square wave, then we will see *tilt* on the top of the square wave. Downward tilt, more commonly called *sag*, indicates LF loss, whilst upward tilt indicates LF boost. This is a very sensitive test of low frequency response, and if it is known that the circuit being measured includes a *single* high-pass filter, but with a cut-off frequency too low to be measured directly with sine waves, then a square wave may be used to infer the sine wave $f_{-3\,dB}$ point. The full derivation of the equation that produced the following table is given in the Appendix, but if we apply a square wave of frequency *f*:

Sag observed using a square wave of frequency '*f*'	**Ratio of applied square wave frequency (*f*) to low frequency cut-off ($f_{-3\ dB}$)**
10%	30
5%	60
1%	300

Most analogue audio oscillators are based on the Wien bridge, and because of amplitude stabilization problems, rarely produce frequencies lower than ≈10 Hz. 10% and 5% square wave sag may be measured relatively easily on an oscilloscope, and can therefore be used to infer sine wave performance at frequencies below 10 Hz.

Another useful test is to apply a high level, high frequency *sine* wave. If at all levels and frequencies, the output is still a sine wave, then the amplifier is likely to be free of slewing distortion. If the output begins to look like a triangular waveform, this is because one, or more, of the stages within the amplifier is unable to charge fully and discharge a shunt capacitance sufficiently quickly. The distortion is known as slewing distortion because the waveform is unable to slew correctly from one voltage to another. The solution is usually to increase the anode current of the offending stage, thereby enabling it to charge and discharge the capacitance.

Random noise

The signals that we have previously considered have been repetitive signals – we could always predict precisely what the voltage level would be at any given time. Besides these *coherent* signals, we shall now consider *noise*.

Noise is all around us, from the sound of waves breaking on a seashore, and the radio noise of stars, to the daily fluctuations of the stock markets. Electrical noise can generally be split into one of two categories. *White* noise, which has constant level with frequency (like white light), and $1/f$ noise whose amplitude is inversely proportional to frequency.

White noise is often known as Johnson or thermal noise, and is caused by the random thermal movement of atoms knocking the free electrons within a conductor. Because it is generated by a thermal mechanism, cooling critical devices reduces noise, and radio telescopes cool their head amplifiers with liquid nitrogen. All resistors produce white noise, and will generate a noise voltage:

$$v_{\text{noise}} = \sqrt{4kTBR}$$

where: k = Boltzmann's constant $\approx 1.381 \times 10^{-23}$ J/K
T = absolute temperature of the conductor = °C + 273.16
B = bandwidth of the following measuring device
R = resistance of the conductor

From this equation, we can see that if we were to cool the conductor to 0 K or −273.16°C, there would be no noise because this would be absolute zero, at which temperature there is no thermal vibration of the atoms to produce noise.

The *bandwidth* of the measuring system is important too, because the noise is proportional to the square root of bandwidth. Bandwidth is the difference between the upper and lower $f_{-3\,dB}$ limits of measurement. It is important to realize that in audio work, the noise measurement bandwidth is always that of the human ear (20 Hz–20 kHz), and although one amplifier might have a wider bandwidth than another, this does not necessarily mean that it will produce any more noise.

In audio, we cannot alter the noise bandwidth or the value of Boltzmann's constant, and reducing the temperature is expensive, so our main weapon in reducing noise is to reduce resistance. We will look into this in more detail in Chapter 7.

1/*f* noise is also known as *flicker* noise or *excess* noise, and it is a particularly insidious form of noise, because it is not predictable. It could almost be called 'imperfection' noise, because it is generally caused by imperfections such as imperfectly clean 'clean rooms' used for making semiconductors or valves, 'dry' soldered joints, poor metal-to-metal contacts in connectors – the list is endless. Semiconductor manufacturers usually specify the 1/*f noise corner* where the 1/*f* noise becomes dominant over white noise for their devices, but equivalent data do not currently exist for valves.

Because noise is random, or *incoherent*, we cannot add noise voltages or currents, but must add noise *powers*, and some initially surprising results emerge. Noise can be considered statistically as a deviation from a mean value. When an opinion poll organization uses as large a sample as possible to reduce error, it is actually *averaging* the noise to find the mean value.

If we parallel '*n*' input devices in a low noise amplifier, the incoherent noise signal begins to cancel, but the wanted signal remains at constant level, resulting in an improvement in signal to noise ratio of $\sqrt{n}$ dB. This technique is feasible for semiconductors where it is possible to make 1000 matched paralleled transistors on a single chip (LM394, MAT-01, etc.), but we are lucky to find a pair of matched triodes in one envelope, let alone more than that!

Active devices

Previously, we have investigated resistors, capacitors and inductors, but these were all *passive* components. We will now look at *active* devices, which can *amplify* a signal. All active devices need a power supply, because amplification is achieved by the source controlling the flow of energy from a power supply into a load via the active device.

We will conclude this chapter by looking briefly at semiconductors. It might seem odd that we should pay any attention at all to semiconductors, but a modern

valve amplifier generally contains rather more semiconductors than valves, so we need some knowledge of these devices to assist the design of the (valve) amplifier.

Conventional current flow and electron flow

When electricity was first investigated, the electron had not been discovered, and so an arbitrary direction for the flow of electricity was assumed. There was a 50/50 chance of guessing correctly, and the early researchers were unlucky. By the time that the mistake was discovered, and it was realized that electrons flowed in the opposite direction to the way that electricity had been thought to flow, it was too late to change the convention.

We are therefore saddled with a conventional current that flows in the *opposite* direction to that of the electrons. Mostly, this is of little consequence, but when we consider the internal workings of the transistor and the valve, we must bear this distinction in mind.

Silicon diodes

Semiconductor devices are made by doping regions of crystalline silicon to form areas known as *N-type* or *P-type*. These regions are permanently charged, and at their junction this charge forms a potential barrier that must be overcome before forward conduction can occur. Reverse polarity strengthens the potential barrier, so no conduction occurs.

The *diode* is a device that allows current to flow in one direction, but not the other. Its most basic use is therefore to *rectify* AC into DC. The arrow head on the diode denotes the direction of conventional current flow, and R_L is the load resistance. See Fig. 1.35.

Practical silicon diodes are not perfect rectifiers, and require a forward *bias* voltage before they will conduct. At room temperature, this bias voltage is

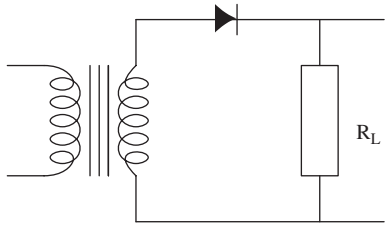


Fig. 1.35 *Use of a diode to rectify AC*

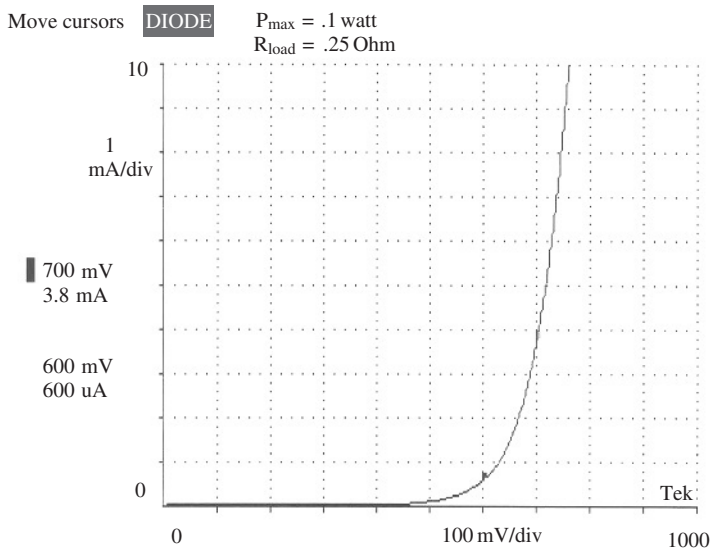


Fig. 1.36 *Silicon diode current against forward bias voltage*

between 0.6 V and 0.7 V but it is dependent on the current passing through the diode. See Fig. 1.36.

This forward bias voltage is always present, so the output voltage is always less than the input voltage by the amount of the *diode drop*. Because there is always a voltage drop across the diode, current flow must create heat, and sufficient heat will melt the silicon. All diodes therefore have a maximum current rating.

In addition, if the reverse voltage is too high, the diode will break down and conduct, and if this reverse current is not limited, the diode will be destroyed.

Unlike valves, the mechanism for conduction through the most common (bipolar junction) type of silicon diode is complex, and results in a charge being temporarily stored within the diode. When the diode is switched off by the external voltage, the charge within the diode is quickly discharged and produces a brief current pulse that can excite external resonances. Fortunately, *Schottky* diodes do not exhibit this phenomenon, and *soft recovery* types are fabricated to minimize it.

Voltage references

The forward voltage drop across a diode junction is determined by the Ebers/Moll equation involving absolute temperature and current, whilst the reverse

breakdown voltage is determined by the physical construction of the individual diode. This means that we can use the diode as a rectifier, or as a voltage reference.

Voltage references are also characterized by their *slope resistance*, which is the Thévenin resistance of the voltage reference when operated correctly. It does not imply that large currents can be drawn, merely that for small current changes in the linear region of operation, the voltage change will be correspondingly small.

Voltage references based on the forward voltage drop are known as *band gap* devices, whilst references based on the reverse breakdown voltage are known as *Zener* diodes. All voltage references should pass only a limited current to avoid destruction, and ideally, a constant current should be passed.

Zener diodes are commonly available in power ratings up to 75 W although the most common rating is 400 mW, and voltage ratings are from 2.7 V to 270 V. Reverse biased junctions produce more noise than forward biased diodes, and so Zeners are noisier than band gap references. In reality, true Zener action occurs at $\leq$5 V, so a 'Zener' diode with a rating $\geq$5 V actually uses the avalanche effect.[2] This is fortuitous, because producing a 6.2 V device requires both effects to be used in proportions that cause the two different mechanisms to cancel the temperature coefficient almost to zero and reduce slope resistance. Slope resistance rises sharply below 6 V, and more gradually above 6 V, but is typically $\approx$10 Ω at 5 mA for a 6.2 V Zener.

Band gap references are often actually a complex integrated circuit, and usually have an output voltage of 1.2 V, but internal amplifiers may increase this to 10 V or more. Because they are complex internally, band gap references tend to be more expensive than Zeners, and we may occasionally need a cheap low noise reference. Light Emitting Diodes (LEDs) and small-signal diodes are operated forward biased, so they are quiet and reasonably cheap.

Diode type	**Typical forward drop @ 10 mA**	**Typical $r_{internal}$ @ 10 mA**
Small-signal silicon diode (1N4148)	0.75 V	6.0 Ω
Infrared LED (950 nm)	1.2 V	5.4 Ω
Cheap red LED	1.7 V	4.3 Ω
Cheap yellow, yellow/green LED	$\approx$2 V	10 Ω
True green LED (525 nm)	3.6 V	30 Ω
Blue LED (426 nm)	3.7 V	26 Ω

Band gap references usually incorporate an internal amplifier, so their output resistance is much lower, typically $\approx 0.2\,\Omega$ or less.

Bipolar Junction Transistors (BJTs)

BJTs are the most common type of transistor; they are available in NPN and PNP types, and can be used to amplify a signal. The name transistor is derived from *trans*ferred res*istor*.

We can imagine an NPN transistor as a sandwich of two thick slabs of N-type material separated by an extremely thin layer of P-type material. The P-type material is the *base*, whilst one of the N-types is the *emitter* and emits electrons, which are then collected by the other N-type, which is known as the *collector*.

If we simply connect the collector to the positive terminal of a battery, and the emitter to the negative, no current will flow because the negatively charged base repels electrons. If we now apply a positive voltage to the base to neutralize this charge, the electrons will no longer be repelled, but because the base is so thin, the attraction of the strongly positively charged collector pulls most of the electrons straight through the base to the collector, and collector current flows.

The base/emitter junction is now a forward biased diode, so it should come as no surprise to learn that 0.7 V is required across the base/emitter junction to cause the transistor to conduct electrons from emitter to collector. Because the base is so thin, and the attraction of the collector is so great, very few electrons emerge from the base as base current, so the ratio of collector current to base current is high. The transistor therefore has current gain, which is sometimes known as β (beta), but more commonly as h_{FE} for DC current gain, or h_{fe} for AC current gain. For all practical purposes, $h_{FE} = h_{fe}$, and the parameter has such wide production tolerances that the distinction is trivial. (h_{FE}: '*h*' signifies the hybrid model, 'F' forward current transfer ratio, and '*E*' that the emitter is the common terminal. Don't you wish you hadn't asked?)

A far more important and predictable parameter is the *transconductance* (previously called *mutual conductance*), *gm*, which is the change in collector current caused by a change in base/emitter voltage:

$$gm = \frac{\Delta I_C}{\Delta V_{BE}}$$

When we look at valves, we will see that we must always measure *gm* at the operating point, perhaps from a graph. For transistors, transconductance is

defined by the Ebers/Moll equation, and for small currents (less than ≈100 mA), *gm* can be estimated for any BJT using:

$$gm \approx 35 I_C$$

The common emitter amplifier

Now that we have a means of predicting the change in collector current caused by a change in V_{BE}, we could connect a resistor R_L in series with the collector and the supply to convert the current change into a voltage change. By Ohm's law:

$$\Delta V_{CE} = \Delta I_C \cdot R_L$$

But $\Delta I_C = gm \cdot \Delta V_{BE}$, so:

$$\Delta V_{CE} = \Delta V_{BE} \cdot gm \cdot R_L$$

$$\frac{\Delta V_{CE}}{\Delta V_{BE}} = A_v = gm \cdot R_L$$

We are now able to find the *v*oltage gain A_v for this circuit, which is known as a common emitter amplifier because the emitter is common to both the input and output circuits or *ports*. Note that the output of the amplifier is *inverted* with respect to the input signal. See Fig. 1.37.

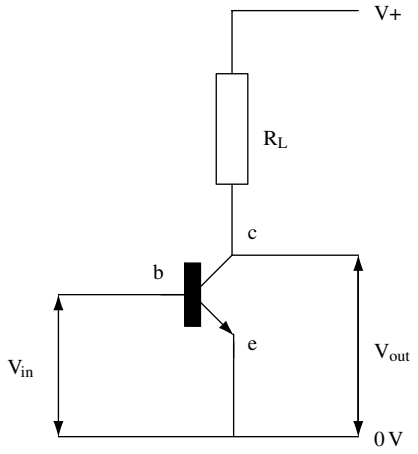


Fig. 1.37 *Common emitter transistor amplifier*

As the circuit stands, it is not very useful, because any input voltage below +0.7 V will not be amplified, so the amplifier therefore creates considerable distortion.

To circumvent this, we assume that the collector voltage is set to half the supply voltage, because this allows the collector to swing an equal voltage both positively and negatively. Because we know the voltage across the collector load R_L, and its resistance, we can use Ohm's law to determine the current through it – which is the same as the transistor collector current. We could then use the relationship between I_B and I_C to set this optimum collector current. Unfortunately, the value of h_{FE} for a transistor is not guaranteed, and varies widely between devices. For a small-signal transistor, it could range from 50 to more than 400. The way around this is to add a resistor in the emitter, and to set a base *voltage*. See Fig. 1.38.

The amplifier now has input and output *coupling* capacitors, and an emitter *decoupling* capacitor. The entire circuit is known as a stabilized common emitter amplifier, and is the basis of most linear transistor circuitry.

Considering DC conditions only: the potential divider chain passes a current that is at least 10 times the expected base current, and therefore sets a fixed voltage at the base of the transistor independent of base current. Because of diode drop, the voltage at the emitter is thus 0.7 V lower. The emitter resistor has a *fixed* voltage across it, and it must therefore pass a fixed current from the emitter. $I_E = I_C - I_B$, but since I_B is so small, $I_E \approx I_C$, and if we have a fixed emitter current, then collector current is also fixed.

The emitter is decoupled to prevent *negative feedback* from reducing the AC gain of the circuit. We will consider negative feedback later in this chapter.

Briefly, if the emitter resistor was not decoupled, then any change in collector current (which is the same as emitter current), would cause the

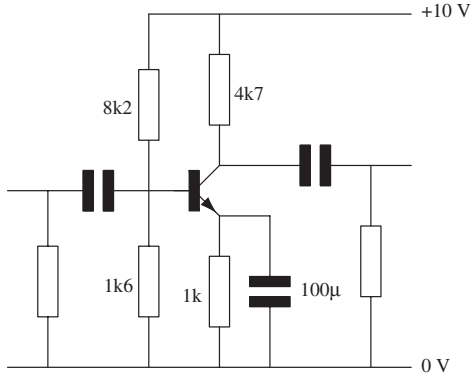


Fig. 1.38 *Stabilized common emitter amplifier*

voltage across the emitter resistor to change with the applied AC. The emitter voltage would change, and V_{be} would effectively be reduced, causing AC gain to fall. The decoupling capacitor is a short circuit to AC, and therefore prevents this reduction of gain. This principle will be repeated in the next chapter when we look at the cathode bypass capacitor in a valve circuit.

Input and output resistances

In a valve amplifier, we will frequently use transistors as part of a bias network, or as part of a power supply, and being able to determine input and output resistances is therefore important. The following AC resistances looking into the transistor do not take account of any external parallel resistance from the viewed terminal to ground.

The output resistance $1/h_{oe}$ looking into the collector is high, typically tens of kΩ at $I_c \approx 1$ mA, which suggests that the transistor would make a good constant current source. (h_{oe}: hybrid model, output admittance, emitter as common terminal.) In theory, if all the collector curves of a bipolar junction transistor are extended negatively, they intersect at a voltage known as the Early voltage. Sadly, the real world is not so tidy, and recent attempts by the author to determine Early voltages resembled the result of dropping a handful of uncooked spaghetti on the graph paper. Nevertheless, the concept of the Early effect is useful because it indicates that $1/h_{oe}$ falls as I_C rises. See Fig. 1.39.

The resistance looking into the emitter is low, $r_e = 1/gm$, so $\approx 20\,\Omega$ is typical. If the base is *not* driven by a source of zero resistance ($R_b \neq 0$), there is an additional series term, and r_e is found from:

$$r_e = \frac{1}{gm} + \frac{R_b}{h_{fe}}$$

R_b is the Thévenin resistance of all the paths to ground *and supply* seen from the base of the transistor. Note that even though we are no longer explicitly including a battery as the supply, the supply is still assumed to have zero output resistance from DC to light frequencies.

Looking into the base, the path to ground is via the base/emitter junction in series with the emitter resistor. If the emitter resistor is not decoupled, then the resistance will be:

$$r_b = h_{fe}\left(\frac{1}{gm} + R_e\right)$$

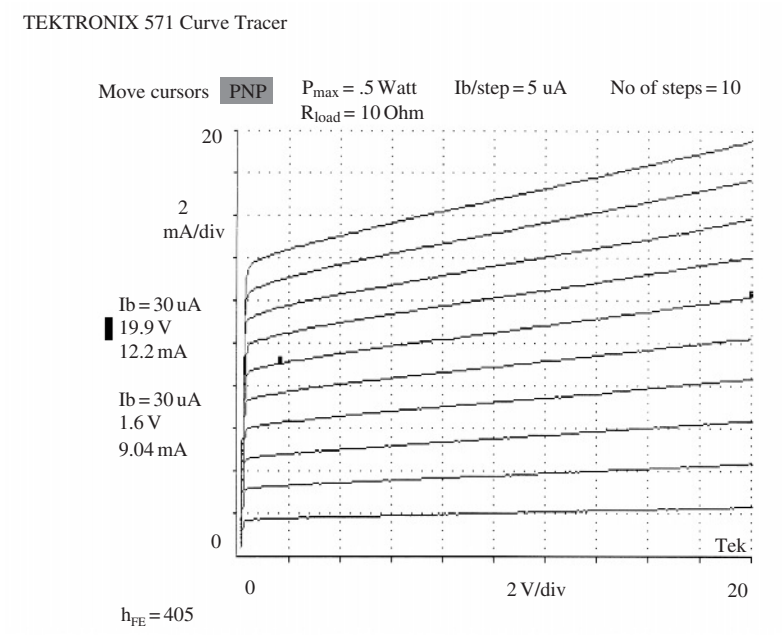


Fig. 1.39 *Collector curves for BC558B PNP transistor*

If the emitter resistor is decoupled, then $R_e = 0$, and the equation reduces to:

$$r_b = \frac{h_{fe}}{gm} = h_{ie}$$

The AC input resistance due purely to the base/emitter junction is often known as h_{ie}, and is generally quite low, $<10\,k\Omega$. (h_{ie}: hybrid model, input resistance, emitter as common terminal.)

If we now consider the effect of the external parallel resistances, we see that the input resistance of the amplifier is low, typically $<5\,k\Omega$. The output resistance seen at the emitter is low, typically $<100\,\Omega$ (even if the source resistance is quite high), and the output resistance at the collector is $\approx R_L$.

The emitter follower

Very occasionally, you will see this amplifier called a *common collector* amplifier, although this phraseology is rare because it does not convey what the circuit is actually doing.

If we reduce the collector load to zero and take our output from the emitter, then we have an amplifier with $A_v \approx 1$. The voltage gain must be ≈ 1, because $V_E = V_B - 0.7\,V$; the emitter *follows* the base voltage, and the amplifier is *non-inverting*. Although $A_v = 1$, the current gain is much greater, and we can calculate input and output resistances using the equations presented for the common emitter amplifier. See Fig. 1.40.

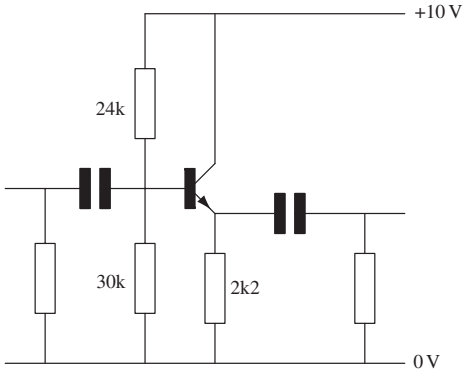


Fig. 1.40 *Emitter follower*

Because of its low output resistance and moderately high input resistance, the emitter follower is often used as a *buffer* to match high impedance circuitry to low resistance loads.

The Darlington pair

Sometimes, even an emitter follower may not have sufficient current gain, and the solution is to use a *Darlington pair*, this is effectively two transistors in cascade, with one forming the emitter load for the other. See Fig. 1.41.

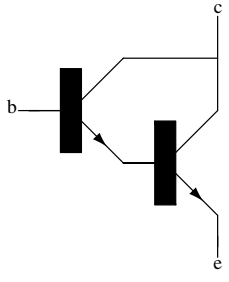


Fig. 1.41 *Darlington pair*

The two transistors form a composite transistor with $V_{BE} = 1.4\,V$, and $h_{FE(total)} = h_{FE1} \times h_{FE2}$. A Darlington pair can replace a single transistor in any configuration if it seems useful. Common uses are in the output stage of a power amplifier and in linear power supplies. Darlingtons can be bought in a single package, but making your own out of two discrete transistors is often cheaper and better.

General observations on BJTs

We mentioned earlier that the BJT could be considered to be a sandwich with the base separating the collector and emitter. We can now develop this model and use it to make some useful generalizations.

As the base becomes thicker, it becomes more and more probable that an electron passing from the emitter to the collector will be captured by the base, and flow out of the base as base current. h_{FE} is therefore inversely proportional to base thickness.

When a transistor passes a high collector current, its base current must be proportionately high. In order for the base not to melt due to this current, the base must be thickened. The thicker base reduces h_{FE}, and required base current must rise yet further, requiring an even thicker base. h_{FE} is therefore inversely proportional to the square of maximum permissible collector current. High current transistors have low h_{FE}.

High voltage transistors must have a thick base in order for the base not to break down under the stress of the voltage that is being insulated from collector to emitter. High voltage transistors have low h_{FE}.

High current transistors must have a large silicon die area in order for the collector not to melt – this large area increases collector/base capacitance. The significance of capacitance in amplifying devices will be made clear when we meet Miller effect in Chapter 2, but for the moment, we can simply say that high current transistors will be *slow*.

We have barely scratched the surface of semiconductor devices and circuits. Other semiconductor circuits will be presented as, and when, they are needed.

Feedback

The feedback equation

Feedback is a process whereby we take a fraction of the output of an amplifier and sum it with the input. If, when we sum it with the input, it causes the gain

of the amplifier to increase, then it is known as *positive* feedback, and this is the basis of oscillators. If it causes the gain to fall, then it is known as *negative* feedback, and this technique is widely used in audio amplifiers.

The description of feedback was deliberately rather vague because there are many ways that feedback can be applied, and they each have differing effects. Before we can look at these effects, we need a few definitions and a simple equation:

$$A = \frac{A_0}{1 + \beta A_0}$$

This is the general feedback equation that defines how the gain of an amplifier will be modified by the application of feedback. β is the *feedback fraction*, and is the proportion of the output that is fed, or looped, back to the input, and A_0 is the gain before feedback.

If βA_0 is very large *and positive* (causing negative feedback – a reduction of gain), then $\beta A_0 \approx \beta A_0 + 1$, and the gain of the amplifier becomes:

$$A = \frac{1}{\beta}$$

A_0 no longer affects A, and the *closed loop* gain of the amplifier is determined solely by the network that provides the feedback signal.

This result is very significant because it implies many things:

- Distortion is produced by variations in gain from one voltage level to another. If open *loop gain* is no longer part of the equation, then small variations in this gain are irrelevant, and the amplifier produces no distortion.
- If the feedback acts to maintain the correct gain *under all circumstances*, then it must change the apparent input and output resistances of the amplifier.
- If the feedback fraction β is set by pure resistors, then the equation for closed loop gain does not contain any term including frequency. Theoretically, the output amplitude is therefore independent of frequency.

In the late 1970s, when cheap gain became readily available, designers became very excited by the possibilities and implications of the feedback equation, and set out to exploit it by designing amplifiers that were thought to have very high levels of feedback. In practice, these amplifiers did *not* have

high levels of feedback at all frequencies and power levels, and it was the *lack* of feedback to linearize these fundamentally flawed circuits that caused their poor sound quality.

Before we explore the expected benefits of feedback, we should therefore examine how the feedback equation could break down.

Practical limitations of the feedback equation

The feedback equation implies improves performance provided that $\beta A_0 >> 1$. If, for any reason, the open loop gain of the amplifier is less than infinite, then βA_0 will *not* be much greater than 1, and the approximation will no longer be true.

Practical amplifiers always have finite gain, moreover, this gain falls with frequency. A practical amplifier will always distort the input signal, and because the distortion reducing ability of negative feedback falls with frequency, the closed loop distortion *must* rise with frequency.

Crossover distortion in Class B amplifiers can be considered to be a reduction of gain as the amplifier traverses the switching point of the transistors or valves. Because of the drastically reduced open loop gain in this region, negative feedback is not very effective at reducing crossover distortion.

Although not explicitly stated in the feedback equation, the phase of the feedback signal is crucially important. If the phase should change by 180°, then the feedback will no longer be negative, but positive, and our amplifier may turn into an oscillator.

We will explore the practical limitations of feedback in Chapter 6, but we should realize that, as with any weapon wielded carelessly, it is possible to shoot oneself in the foot . . .

Feedback terminology and input and output impedances

The way in which the feedback is *derived* affects the output impedance, whereas the way that it is *applied* will affect input impedance. If we make a *parallel*, or *shunt* connection, then we are dealing with a *voltage*, but if we make a *series* connection, we are dealing with a *current*. Feedback confined to one stage is known as *local* feedback, whereas feedback over a number of stages is known as *global*. We may have more than one global feedback loop, with one loop enclosed by another, in which case, the loops are said to be *nested*.

Now that we have defined how the feedback is connected, we can state that for *negative* feedback, voltage feedback *reduces* impedances, and current feedback *increases* impedances.

As an example, we can now combine these terms to describe the global negative feedback loop of a power amplifier as being parallel derived, series applied, but we could equally well describe it as being voltage derived, current applied. Almost all power amplifiers use this particular feedback strategy because it ensures:

- Low output impedance (needed to damp electromagnetic loudspeakers correctly).
- High input impedance (to avoid excessive loading of pre-amplifiers).
- Non-inverting gain.

If we have *positive* feedback, all effects are reversed, so voltage feedback increases impedances whilst current feedback reduces impedances. Using a combination of positive and negative feedback it is possible to make a power amplifier with zero, or even negative, output impedance.

Unfortunately, amplifiers with negative output impedance are liable to oscillate, because they reduce the total series damping resistance of the external (invariably resonant) load to zero. Nevertheless some early valve power amplifiers had the facility to adjust output impedance through zero to a negative value, in an effort to improve the bass performance of the accompanying loudspeaker. Occasionally, this idea resurfaces, but it is really only of use for the dedicated amplifiers in a loudspeaker system with an active crossover. (Active, or low level, crossovers are used *before* the power amplifiers, so that each drive unit has a dedicated power amplifier. Although this scheme may seem profligate with expensive power amplifiers, it has much to recommend it.)

Although we can describe the various modes of feedback, and their effect on impedances, we now need to quantify this effect. Impedances are changed by the ratio of the *feedback factor*:

$$\text{feedback factor} = (1 + \beta A_0)$$

This is also the factor by which the gain of the amplifier has been reduced, and is often expressed in dBs. An amplifier with 20 dB of global feedback has had its total gain reduced by a factor of 20 dB, and if this was a conventional power amplifier with shunt derived feedback, its output impedance would be reduced by a factor of 10.

The operational amplifier

Conventionally, when we think of computers, we think of *digital* computers but once upon a time there were also *analogue* computers, which were

hardwired to model complex differential equations, such as those required for calculating the ballistics of shells. These analogue computers used a basic building block that became known as the *operational amplifier*. This was a small device for its time (smaller than a brick) and used valve circuits that operated from ±300 V supplies. With suitable external components, these operational amplifiers could be made to perform the mathematical operations of inversion, summation, multiplication, integration and differentiation.

Thankfully, the valve analogue computer is no longer with us, but the term 'operational amplifier', usually shortened to *op-amp*, is still with us, even if it now refers to eight-legged silicon beetles.

Op-amps attempt to define the performance of the final amplifier *purely* by feedback, and to do this successfully, the op-amp must have enormous gain; 120 dB gain at DC is not uncommon, although AC gain invariably falls with frequency.

In the following discussions, we will make two fundamental assumptions about op-amps:

- Gain is infinite.
- Input resistance is infinite.

These assumptions specify ideal op-amps. Real world op-amps have limitations, and will *not* achieve this ideal at all frequencies, voltage levels etc. Provided that we remember this important caveat *at all times*, we will not run into trouble.

It is habitual in op-amp circuit diagrams to omit the (usually ±15 V) power supply lines to the op-amp to aid clarity, nevertheless, the op-amp will still need power!

The invertor, and virtual earth adder

The op-amp invertor has parallel derived, parallel applied, feedback, and since op-amp gain is infinite, both the point of derivation and application must have zero resistance, the amplifier has zero output resistance. See Fig. 1.42.

The inverting input of the op-amp also has zero resistance to earth (0 V) because of the feedback. Since the gain of the op-amp is infinite, if the non-inverting input is at earth potential, then the inverting input must also be at earth potential. In this configuration, the inverting input is therefore known as a *virtual earth*.

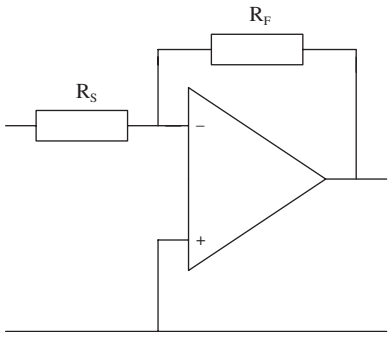


Fig. 1.42 *Inverting amplifier*

Although it is a virtual earth by virtue of feedback, the inverting input of the op-amp itself has infinite resistance, and *no signal current flows into the op-amp*. Input signal current from R_S can therefore only flow to ground via R_F and the zero output resistance of the op-amp. The signal currents in R_S and R_F are therefore equal, and using Ohm's law:

$$\frac{V_S}{R_S} = \frac{V_F}{R_F}$$

Since the inverting input is a virtual earth, $V_S = V_{in}$, and $V_F = -V_{out}$ (the op-amp inverts), and the voltage gain of the amplifier is:

$$A_v = \frac{V_{out}}{V_{in}} = -\frac{R_F}{R_S}$$

Note that this amplifier can achieve $A_v < 1$, and can *attenuate* the input signal. This *is* useful because it provides an attenuated output from an almost zero source resistance, whereas a potential divider would have significant output resistance. The minus sign reminds us that the amplifier is inverting.

Because the inverting node of the amplifier is a virtual ground, input resistance is equal to R_S.

When we analysed this amplifier, we considered the input signal current. There is no reason why this input current should come from only one source via one resistor, and we can sum currents at the inverting node in accordance with Kirchhoff's law. See Fig. 1.43.

This circuit is known as the *virtual earth adder*, and is most useful in bias servo circuits where we may need to add a number of correction signals. Voltage gain for

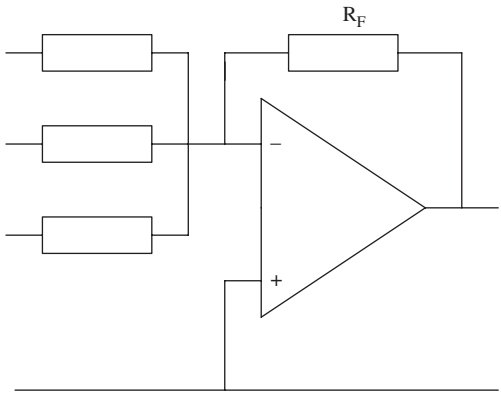


Fig. 1.43 *Virtual earth adder*

each input may be determined using the invertor equation. When multiple inputs are driven, it is often best to determine the output voltage by summing the signal currents using Kirchhoff's law, finding the resultant current in R_F, and using Ohm's law to determine the output voltage. This is not as tedious as it sounds.

The non-inverting amplifier and voltage follower

Frequently, we may need a non-inverting amplifier. See Fig. 1.44.

In this configuration, we still have R_F and R_S, but the amplifier has been turned upside down, and the far end of R_S is now connected directly to ground. The amplifier has parallel derived, series applied, feedback. Output resistance is

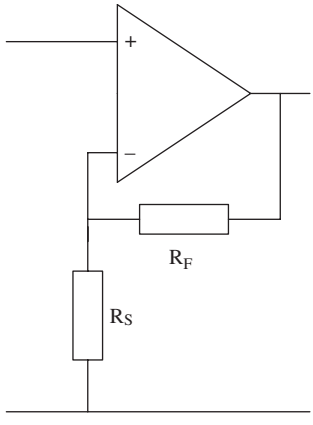


Fig. 1.44 *Non-inverting amplifier*

therefore zero, whilst input resistance is infinite. R_F and R_S now form a potential divider across the output of the amplifier, and the voltage at the inverting input is:

$$V_{(\text{inverting input})} = \frac{R_S}{R_S + R_F} \cdot V_{out}$$

Since the op-amp has infinite gain, the voltage at the inverting input is equal to that at the non-inverting input, which is V_{in}, so the gain of the amplifier is:

$$\frac{V_{out}}{V_{in}} = \frac{R_S + R_F}{R_S}$$

If we now reduce the value of R_F to zero, and discard R_S, the gain reduces to 1, and the amplifier is known as the *voltage follower*. See Fig. 1.45.

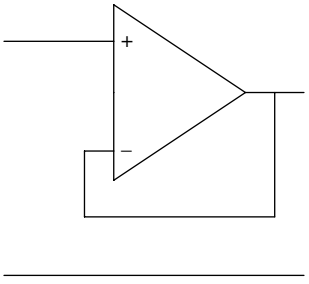


Fig. 1.45 *Voltage follower*

The voltage follower is an excellent *buffer* stage for isolating high impedance circuits from low impedance loads, and is far superior to the single transistor emitter follower that we saw earlier. If we need more current than the op-amp can provide, we can add an emitter follower, or even a Darlington pair, to the output, and enclose this within the feedback loop. See Fig. 1.46.

The integrator

This circuit is essentially an invertor with R_F paralleled by C_F, but if R_F is omitted, it becomes an *integrator* and measures *charge*. The circuit may be considered to be a low-pass filter, in which case, the cut-off frequency of the filter is:

$$f_{-3\,\text{dB}} = \frac{1}{2\pi C_F R_F}$$

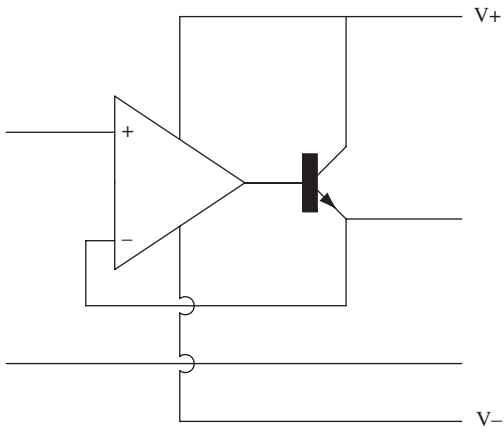


Fig. 1.46 *Addition of emitter follower to increase output current*

We can now see that if $R_F = \infty$, the cut-off frequency = 0, and the amplifier has infinite gain at DC. Any DC offsets will gradually accumulate sufficient charge on the capacitor to cause the output of the op-amp to rise to maximum output voltage ($\pm V_{supply}$), and the op-amp is then said to be *saturated*. For this reason, practical integrators are usually enclosed by a servo control loop. The servo error signal charges C_F, but the loop aims to minimize the error, and so the voltage across C_F tends to zero.

Again, we could sum several input currents if we wished, and one possible use of this circuit would be for monitoring the voltages across the cathode resistors of several output valves, because the integrating function would remove any audio signal and provide a DC output voltage proportional to the *total* output valve DC current.

DC offsets

We briefly mentioned DC offsets when considering the integrator. Op-amps are not magical, they contain real transistors, and base or leakage current, known as *bias current*, will flow out of the inputs. The bias currents will not be perfectly matched, and the imbalance is known as *offset current*. The input transistors will not be perfectly matched for voltage either, so there will be an *offset voltage* between the inputs. These imperfections are detailed in the manufacturer's data sheets, and should be investigated during circuit design.

References

1 Kaye and Laby (1995) *Tables of Physical and Chemical Constants*, 16th ed. Longman.
2 Greg Parker (1994) *Introductory Semiconductor Device Physics*, Prentice Hall.

Further reading

Duncan, Ben (1997) *High Performance Audio Power Amplifiers*. Newnes.
Hartley Jones, Martin (1995) *A Practical Introduction to Electronic Circuits*, 3rd ed. Cambridge University Press.
Horowitz, P. W. and Hill, W. (1989) *The Art of Electronics*, 2nd ed. Cambridge University Press.
Linsley Hood, John (1998) *The Art of Linear Electronics*, 2nd ed. Newnes.
Self, Douglas (2000) *Audio Power Amplifier Design Handbook*, 2nd ed. Newnes.

2

Basic building blocks

In this chapter we will look mainly at the triode valve, how to choose operating conditions, and what effect these choices have on the AC performance of the stage. The analysis will use a combination of graphical and algebraic techniques, which has the advantage of being quick to use, and the results of the theory agree well with practice. This last point might seem to be an obvious requirement, but it is one that is sometimes overlooked.

The common cathode triode amplifier

The most common use of a valve is amplification. Therefore we need to know how to configure and bias the valve so that it can amplify in a linear manner and minimize distortion. We will begin by investigating the *anode characteristics* of an ECC83/12AX7. See Fig. 2.1.

The anode characteristics are the most useful set of curves for a valve, and the plot shows anode current I_a against anode voltage V_a, for differing values of grid to cathode voltage (V_{gk}). The first point to note is that valves operate at high voltages (typically a factor of ten greater than transistor circuits), and quite low currents. The second point is that if there is no *bias* voltage ($V_{gk} = 0$), then a large anode current flows. This is known as the *space-charge limited* condition, and means that the flow of current is limited only by the number of electrons that can be released from the cathode. In contrast to the bipolar junction transistor, we have to turn the triode off, rather than on, in order to bias it correctly.

The basic amplifier stage has an *anode load* resistor R_L, connected between the anode and the *HT* supply (this is the historical phraseology, and stands for High Tension). See Fig. 2.2.

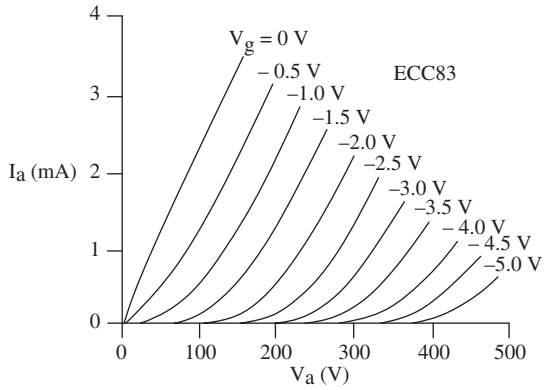


Fig. 2.1 *Triode anode characteristics*

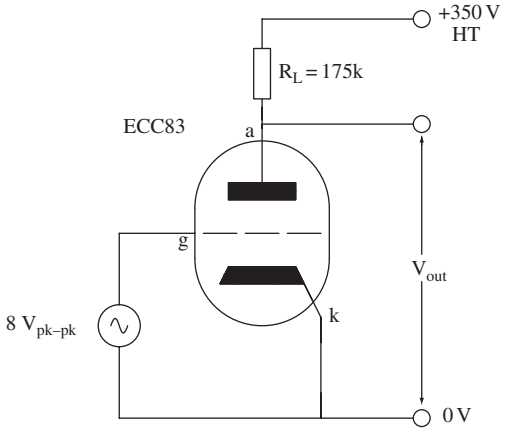


Fig. 2.2 *Common cathode amplifier*

The HT supply is assumed to have zero output resistance at all frequencies from DC to light (you may wish to consider whether this is in fact the case in a practical amplifier). By applying our input voltage between the grid and cathode, we modulate V_{gk}, and thereby control anode conditions. This is why this grid is often known as the *control grid* in multi-grid valves such as tetrodes and pentodes.

We will now use the technique of *loadlines* to link the amplifier circuit to the anode characteristics, and to extract useful information from them.

Using Ohm's law, it is apparent that if there is no current flowing through the resistor (and therefore the valve), there must be no voltage across the resistor. If there is no voltage across the resistor, then all of the HT must be

across the valve, so we could mark that as a point on the graph of the anode characteristics ($V_a = \text{HT} = 350\,\text{V}$, $I_a = I_R = 0$). Similarly, we can argue that if there is no voltage across the valve, then the HT must all be across the resistor; we can calculate the current through the resistor, and therefore the valve. In this case, $R_L = 175\,\text{k}\Omega$, $\text{HT} = 350\,\text{V}$, so the anode current $I_a = 2\,\text{mA}$, and we can plot this point too.

Because Ohm's law is an equation that describes a straight line, if we know two points, we have completely defined that straight line. This means that we can now draw a straight line between our two plotted points, as shown. See Fig. 2.3.

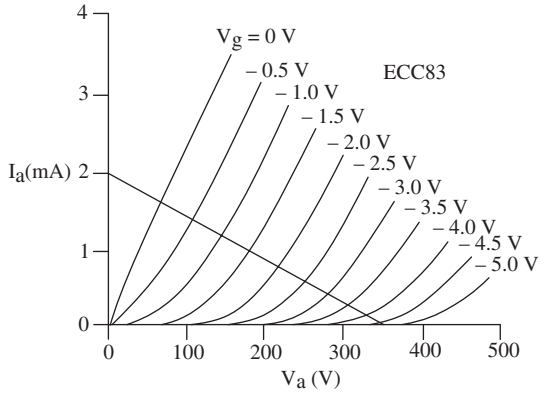


Fig. 2.3 *The loadline*

This is our *loadline*. This is perhaps the single most useful piece of analysis that can be performed on a valve stage. We have defined the anode current for any anode voltage, using an HT of 350 V and anode load of 175 kΩ. If we want to change our anode load, or HT, we must recalculate and redraw our loadline.

If we look along the loadline, we see that it is intersected at various points by the I_a/V_a curves for differing values of V_{gk}. What this means is that those differing values of V_{gk} will cause predictable changes in anode voltage, and we could therefore calculate the *gain* of the stage.

Let us suppose that we apply an $8\,\text{V}_{\text{pk–pk}}$ sine wave to the input of the stage. If we start from 0 V, and look to see where the 0 V grid bias line intersects the loadline, this occurs at $V_a = 72\,\text{V}$. We then let the sine wave swing negative to −4 V, and see that it results in $V_a = 332\,\text{V}$. For an applied voltage of −4 V on the input, we produced a positive change of voltage at the anode of 260 V. The amplifier *inverts*. Since gain is defined as the ratio of

output voltage to input voltage, we have just produced an amplifier with a gain of −65 (the minus sign merely reminds us that this is an inverting amplifier).

Unfortunately, it isn't very linear. If we now allow the input sine wave to continue rising past 0 V, we soon find that the anode voltage is unable to fall any further, and so the output signal no longer looks like the input signal.

We must choose a *bias* or *operating point* at which we will set the *quiescent* (no signal) conditions such that the stage can accommodate both negative and positive excursions of the signal without gross distortion.

Limitations on choice of the operating point

Not only did the previous circuit distort, but the anode DC voltage was superimposed on the output signal, so we add a capacitor and a resistor at the output to block the DC. See Fig. 2.4.

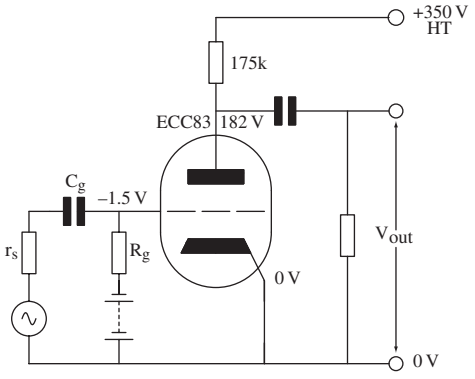


Fig. 2.4 *Grid bias using battery*

The valve is biased by superimposing a bias voltage onto the grid via R_g which prevents the battery from short circuiting the oscillator. C_g is the *coupling capacitor* that prevents the oscillator from shorting the battery, and r_s is the output resistance of the oscillator.

Returning to the loadline, we find that as V_a rises, the grid curves become progressively bunched together, which indicates non-linearity, and is particularly severe when V_a is close to the HT voltage. This region is known as *cut-off* (because the flow of current is being cut off). Operation near cut-off is not advisable if good linearity is required, although we will meet this mode of operation later when looking at some power stages.

Moving in the opposite direction along the loadline, we turn the valve on harder and harder, until finally there is no voltage across it. This is extreme, however, and we will encounter the problem of *positive grid current* long before that. As we make the grid less and less negative, there comes a point when the electrons leaving the cathode are no longer repelled and controlled by the grid, but are actually attracted to the grid and flow out through the grid to ground. This causes the input resistance of the valve, which could previously be regarded as infinite, to fall to a value sufficiently low that it begins to load the oscillator's (non-zero) output resistance. Because this attenuation only happens on the positive *peaks* of the input waveform, this causes distortion of the *input* signal, even though the valve accurately amplifies the grid voltage. The onset of grid current varies with valve type (but is generally around $-1\,V$) and is usually specified on the valve data sheet. For instance, Mullard specifies V_{gk} max. ($I_g = +0.3\,\mu A$) as $-0.9\,V$ for the ECC83 used in our example.

If we have a voltage across the valve, and a current flowing through it, we must be dissipating power within that valve, and there will be a limit beyond which we are in danger of melting the internal structure of the valve. This is known as the *maximum anode dissipation* and is given on the data sheet as being 1 W for the ECC83. For power valves, the curve that corresponds to this is often drawn on the anode characteristic curves, but, if we wish, we can easily add it ourselves. All we need do is to calculate the current drawn for 1 W at 0 V, 50 V, 100 V, 150 V . . . We plot these results on the graph, and draw a curve through the points to form a hyperbola.

The valve data sheet also specifies two more, interlinked, restrictions on the choice of bias point, maximum V_a and maximum $V_{a(b)}$. Maximum V_a is the maximum DC voltage at which the anode may be continuously operated, whereas $V_{a(b)}$ is the maximum voltage to which the anode may be allowed to swing under signal or cold conditions, and is effectively the maximum allowable HT voltage for that valve. Ignoring these limits usually results in premature destruction of the valve to the accompaniment of blue flashes and bangs as residual gas in the valve is ionized and breaks down. This in itself may not cause irreversible damage, but if a path is formed between the anode and the control grid, then a large anode current will flow, and this may damage the valve permanently. You have been warned.

The final limitation is the maximum allowable cathode current $I_{k(max.)}$. One of the other limitations usually comes into effect first, but input stages may minimize V_a and maximize I_a in order to maximize *gm* and minimize noise, so $I_{k(max.)}$ should be checked if possible. (Neither Mullard nor Brimar specified $I_{k(max.)}$ for the ECC83.)

We can now draw these limitations onto the anode characteristics, and choose our operating point from within the clear area. See Fig. 2.5.

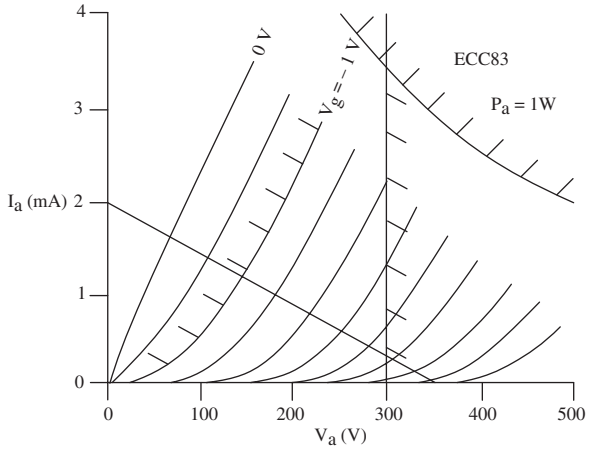


Fig. 2.5 *Determination of safe operating area*

Conditions at the operating point

Although the choice of operating point has now been considerably restricted, we can still optimize various aspects of performance.

In general, there are two main (and usually conflicting) factors: maximum voltage swing and linearity. If we want to bias for maximum voltage swing, then we would set the bias point at $V_a = 225\,V$, to allow the anode to swing up to 300 V and down to 150 V; this would be done by setting the grid bias to $-2.1\,V$.

However, it might be that we are interested more in linearity than in maximum swing. Triodes produce mainly second harmonic distortion, which is generated by the amplifier having unequal gain on the positive half cycle of the waveform compared to the negative half cycle, and the distortion is directly proportional to amplitude. To maximize linearity, we should look for an operating point where the distance to the first grid line either side of the operating point is, as nearly as possible, equal. In this case we might bias the anode voltage to 182 V by applying $-1.5\,V$ to the grid.

Supposing that we have chosen the linearity approach, we will now want to determine the *dynamic* or *AC* conditions of the stage to see if they satisfy our needs.

The first, and most obvious, parameter to determine is the voltage amplification (A_v), or gain, of the stage. We do this by looking an equal distance

either side of the operating point to the first intersection with a grid line, noting the anode voltage. Referring to Fig. 2.5, if we move from the operating point to the right, we meet the 2 V grid line, which intersects at a voltage of 220 V, similarly, the 1 V line intersects at 148 V.

Amplification (A_v) = change in anode voltage/change in grid voltage

$$A_v = \frac{\Delta V_a}{\Delta V_g} = \frac{220 - 148}{1 - 2} = -72$$

The minus sign reminds us that the amplifier is inverting, but you will usually find this dropped, in the interests of clarity, since most stages invert, and the absolute polarity of any particular stage is often of little consequence.

The next important factor is the maximum undistorted voltage swing. Again, we look symmetrically either side of the operating point, but this time we look for the first limiting value. In this instance, we look to the left and see that at 148 V we are approaching positive grid current. This would not matter if our source had zero output resistance, but this is unlikely to be the case, and so we must regard this as a limit. If we look to the right, we find that there is no practical limit until V_a = HT. Unfortunately, whilst this means that the valve can swing a large voltage positively, it cannot swing as far negatively. It is the *first* limit to be reached that is important. We can now see that the maximum undistorted peak to peak swing at the output is double that of the distance from the bias point to the first limit. In this example this corresponds to 72 V_{pk-pk}, but remembering that AC signals are specified as the RMS value of a sine wave, we should divide this figure by a factor of $2\sqrt{2}$, which results in a value of 25 V_{RMS} as the maximum undistorted sine wave output, which is perhaps not so impressive.

It may be that this value of maximum output swing is insufficient, so we would go back and reselect our operating point. If we are still unable to achieve a satisfactory value, then we may need to choose a different value of R_L, HT, or both. It will now be apparent that designing valve stages requires a pencil, a clear ruler, an eraser, and a lot of photocopied anode characteristics curves.

Assuming that the stage looks promising so far, the next important parameter is the output resistance. A triode can be modelled as a voltage source coupled through a series resistance, known as the *anode resistance*, r_a. Remember that because this is an AC, or dynamic, parameter, it is given a lower case letter, and is quite distinct from R_L, the anode load. This dynamic anode resistance is then in parallel with the anode load to form the output resistance r_{out}. See Fig. 2.6.

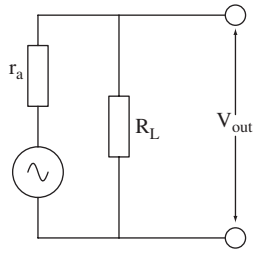


Fig. 2.6 *Thévenin equivalent of triode anode circuit*

It should be noted that the value of gain predicted for the stage by the loadline *already* includes the attenuation caused by the potential divider formed by r_a and R_L.

To find r_a, we return to the anode characteristics and draw a tangent to the curve where it touches the operating point. What we are aiming to do is to measure the gradient of the curve at that point. This is not as difficult as it sounds. A true tangent will touch, or intersect, the curve at only *one* point, and only at the correct point. A good quality transparent ruler is ideal for this purpose. Having positioned the ruler correctly, we draw a line that reaches the edges of the graduations on the data sheet, and read off the values at these points. The purpose of this is to make the resulting triangle, from which we take our figures, as large as possible in order to minimize errors. See Fig. 2.7.

The anode resistance r_a can now be calculated from:

$$r_a = \frac{\Delta V_a}{\Delta I_a} = \frac{V_2 - V_1}{I_2 - I_1} = \frac{382 - 121}{4 - 0} = 65\,\text{k}\Omega$$

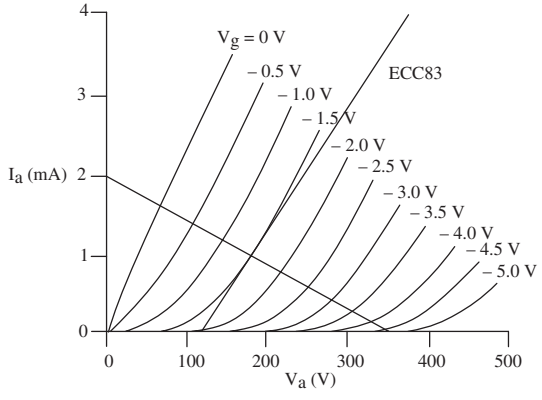


Fig. 2.7 *Determination of dynamic anode resistance r_a*

You will note that the units of mA were used directly in the equation, resulting in an answer in kΩ; this is conventional practice, and saves time. The output resistance r_{out} is simply r_a in parallel with R_L, which results in a value of 47 kΩ. This is quite a high value of output resistance, and is a consequence of using a high μ (mu) valve, as they tend also to have a high value of r_a in operation.

Dynamic, or AC, parameters

So far, we have analysed the behaviour of the valve graphically, but this is not the only method. There are three AC parameters that define completely the characteristics of a valve, *provided that they are evaluated at the operating point*. The importance of this last point is sometimes overlooked.

These parameters are:

μ (mu) = amplification factor (no units)
gm = mutual conductance (usually mA/V)
r_a = anode resistance (kΩ, Ω)
The amplification factor is defined by:

The amplification factor (μ) *of a valve is ratio of the change in anode voltage* ΔV_a *to the change in grid voltage* ΔV_g, *with anode current held constant.*

$$\mu = \frac{\Delta V_a}{\Delta V_g}$$

In a more digestible form, it is the maximum possible voltage amplification of the valve, and can only be achieved if $R_L = \infty$. In practice, we rarely achieve a gain as high as this.

Valves are frequently classified by μ as follows:

Low μ :	<8	$(6080 = 2, 12B4A = 6.5)$
Medium μ :	$8 - 30$	(Type $76 = 13.8$, $ECC82 = 18, 6SN7 = 20$)
High μ :	>30	($ECC81 = 65, 6SL7 = 70$, $ECC83 = 100, ECC807 = 150$, $WE416 = 250, PD500 = 1050$)

We can measure μ at the operating point by drawing a horizontal line through the operating point, which is equivalent to $R_L = \infty$, and calculating the gain as before, by noting the intersections with the grid curves. See Fig. 2.8.

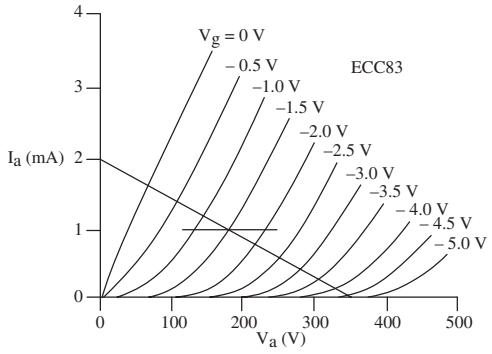


Fig. 2.8 *Determination of μ*

$$\mu = \frac{233 - 133}{2 - 1} = 100$$

Note that it is usual to ignore the signs of the individual voltages measured in equations like this.

Rather than using the loadline, we can use a formula to determine the voltage gain A_v of the amplifier stage:

$$A_v = \mu \cdot \frac{R_L}{R_L + r_a} = 100 \cdot \frac{175}{175 + 65} = 73$$

Which is in good agreement with the value predicted by the loadline ($A_v = 72$). You will find that μ is one of the more stable valve parameters, and varies little with anode current (a fact that will be exploited later). However, this method is not ideal, since the accuracy of the final answer is dependent on how accurately you can draw tangents. It is, however, thoroughly recommended as a check on the general accuracy of your predictions.

The second valve parameter, mutual conductance, is defined by:

The mutual conductance gm of a valve is the ratio of the change in anode current ΔI_a to the change in grid voltage ΔV_g, with anode voltage held constant.

$$gm = \frac{\Delta I_a}{\Delta V_g}$$

Finally, we can define r_a:

The anode resistance r_a of a valve is the ratio of the change in anode voltage ΔV_a to the change in anode current ΔI_a, with grid voltage held constant.

$$r_a = \frac{\Delta V_a}{\Delta I_a}$$

There is a very useful equation which links these three parameters together:

$$gm = \frac{\mu}{r_a}$$

Obviously, we can rearrange this equation as necessary to find the third parameter if we know the other two, but it is specified this way round, because although we can always predict μ and r_a reasonably accurately from the anode characteristics, we cannot directly predict an accurate value for *gm* (yet traditional valve testers very rarely measure any parameter other than *gm*). See Fig. 2.9.

In theory, to find *gm*, we simply draw a vertical line through the operating point (hold V_a constant), and measure the change in anode current. However, it is immediately apparent that the change from 1.5 V to 1 V is considerably greater than the change from 1.5 V to 2 V, and taking the *average* value from 1 V to 2 V does not give an accurate figure of *gm* at the operating point.

$$gm = \frac{1.72 - 0.3}{1 - 2} = 1.42\,\text{mA/V}$$

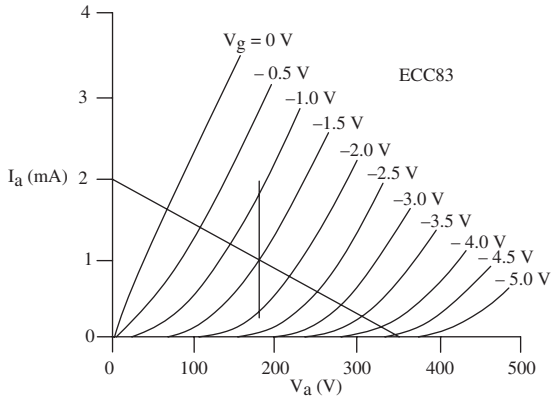


Fig. 2.9 *Determination of gm*

Using the equation, however, with accurate values of μ and r_a (they must be, because they agreed well with the loadline prediction of gain), we find:

$$gm = \frac{\mu}{r_a} = \frac{100}{65} = 1.54\,\text{mA/V}$$

Which means that the previous value was almost 8% low. We will return to use *gm* later.

A parameter that is very occasionally mentioned is *perveance*, which is the ratio of the space-charge limited anode current to the three halves power of anode voltage (Child's law):

$$G = \frac{i_{a\ (\text{space-charge limited})}}{V_a^{\frac{3}{2}}}$$

The practical significance of perveance is that a high perveance valve requires less anode voltage for a given anode current. Additionally, high perveance valves such as the 5687 can often swing their anodes closer to 0 V, increasing voltage swing and efficiency.

Cathode bias

Now that we have chosen our operating point and evaluated the dynamic characteristics of our amplifier stage, we need to look at practical ways of implementing the stage. Whilst we could bias the stage using a battery, it is

inconvenient to disassemble the amplifier just to change a battery. However, lithium batteries having a shelf life of ten years are now readily available, so battery replacement could perhaps be less frequent than valve replacement.

Another way of providing *grid bias* would be to have a subsidiary negative power supply, and use potential dividers to determine the bias to individual valves. This is frequently done on power stages, but could cause noise and stability problems with small-signal stages.

An alternative method is to insert a *cathode bias* resistor between the cathode and ground, and connect the grid to ground via a *grid leak* resistor. Conveniently, the grid is now at 0 V, so we no longer need an input coupling capacitor. See Fig. 2.10.

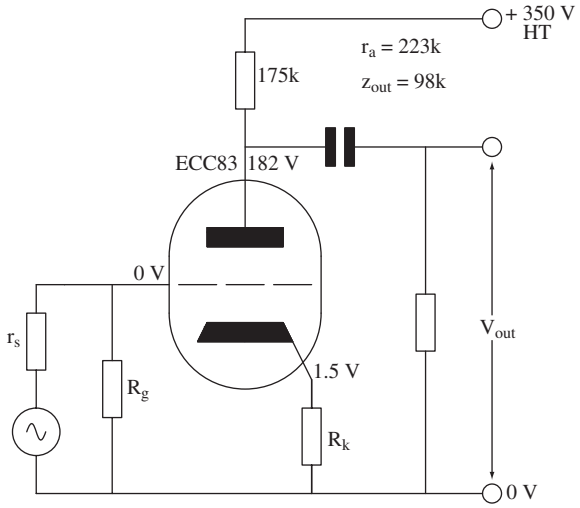


Fig. 2.10 *Cathode bias*

To understand the operation of this stage, we will assume a perfect valve that does not pass grid current even if $V_{gk} = 0$.

Initially there is no current flowing through the valve. If this is the case, there will be no voltage drop across the cathode bias resistor, and the cathode will be at 0 V. The grid is tied to 0 V, so V_{gk} must be 0 V. This will cause the valve to conduct heavily, but as it does so, the anode current (which in a triode is equal to the cathode current) flows through the cathode bias resistor, causing a voltage drop across it. This voltage drop causes the cathode voltage to rise, V_{gk} falls, and an equilibrium anode current is reached.

We know our operating point, therefore we know anode, and hence, cathode current. We know what value of V_{gk} we need. If the grid is at 0 V, then the cathode must be at $+V_{gk}$. If we know the voltage across, and the current through, an unknown resistor, then it is a simple matter to apply Ohm's law and find the value of that resistor. In our example we chose to place our operating point at 182 V. We could read off the anode current directly, but it is more accurate in this instance to calculate the current using Ohm's law. (This is because we can read off the value of V_a with greater accuracy.)

$$I_a = \frac{HT - V_a}{R_L} = \frac{350 - 182}{175} = 0.96\,\text{mA}$$

We know that the cathode voltage is 1.5 V, so the cathode bias resistor will be:

$$R_k = \frac{1.5}{0.96} = 1.56\,\text{k}\Omega$$

Again, note that the equation directly used mA, resulting in a resistance in kΩ.

The effect on AC conditions of an unbypassed cathode bias resistor

Whilst the cathode bias resistor *stabilized* and set the DC conditions of the stage, it did so by means of negative feedback, so we should expect it to affect the AC conditions such as gain and output resistance. We can use the universal feedback equation to determine the effect it will have.

$$A_{fbk} = \frac{A_0}{1 + \beta \cdot A_0}$$

The feedback fraction β in this case is the ratio R_k/R_L, so:

$$A_{fbk} = \frac{72}{1 + \frac{1.56}{175} \cdot 72} = 44$$

The gain has been considerably reduced. The feedback is series derived, and series applied, so it raises the input and output resistances. Since the input resistance of a valve is virtually infinite anyway, this won't be affected, but the anode resistance r_a will be raised.

Although the feedback equation is very handy for quickly determining the new gain, it is not quite so easily used for finding the new r_a.

Looking down through the anode, the only path to ground is the cathode, via the anode resistance r_a. Since, in this direction, resistances are *multiplied* by $(\mu + 1)$, we see an effective anode resistance of:

$$r'_a = r_a + (\mu + 1) \cdot R_k = 65 + (100 + 1) \cdot 1.56 = 223\,k\Omega$$

The value of r_a rises from 65k to 223k. In parallel with R_L, this gives a new output resistance of 98k, as opposed to 47k. Incidentally, there is no reason why we should not calculate the new r_a *first* and use that new value in the standard gain formula to determine the new gain:

$$A_v = \mu \cdot \frac{r'_a}{R_L + r'_a} = 100 \cdot \frac{175}{175 + 223} = 44$$

It is most important to appreciate that the feedback affected only the valve's internal r_a. The anode load resistor R_L was external to the feedback, and therefore not affected.

Having evaluated the new values of gain and output resistance, we may decide that they are no longer satisfactory. We could either choose a new value of R_L, and try a new operating point, or we might even choose a new valve. However, there is another avenue open to us.

The cathode decoupling capacitor

The addition of the cathode bias resistor caused negative feedback, and reduced gain. This may not always be desirable, so we will now consider how to prevent this feedback.

Because the output signal is derived from changing I_a through R_L, and I_a also flows through R_k, we must also develop a signal voltage across R_k. The signal voltage across R_k is in phase with the input signal, but because the valve responds to changes in V_{gk}, which is the *difference* between V_g and V_k, we have effectively reduced the available driving voltage to the valve.

In order to restore full gain, we must suppress the feedback voltage produced at the cathode with a *decoupling* or *bypass* capacitor. The capacitor should be of sufficiently low reactance that it is a short circuit at all AC frequencies of interest. In conjunction with the output resistance at the cathode, this forms a local low-pass filter. See Fig. 2.11.

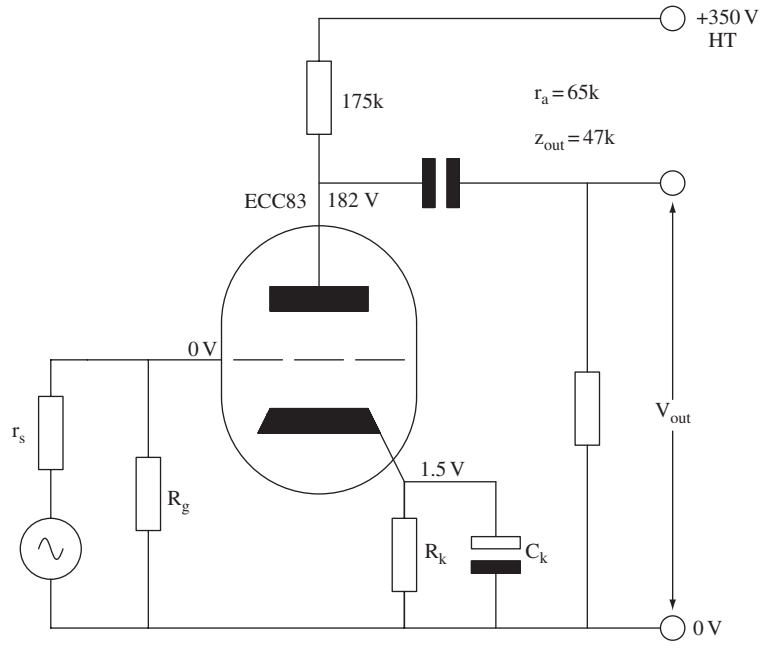


Fig. 2.11 *Cathode decoupling*

We now need to know what resistance the capacitor 'sees' from its positive terminal to ground. Clearly it sees the resistor R_k, but it also sees the cathode resistance of the valve. The resistance looking into the cathode is:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

We can see the HT supply (AC ground) through the series resistance of r_a and R_L, but this is divided by the factor $(\mu + 1)$ due to the internal voltage gain of the stage. If we now put some numbers into the equation, we have:

$$r_k = \frac{175 + 65}{100 + 1} = 2.38\,\text{k}\Omega$$

In parallel with the 1.56 kΩ cathode bias resistor this gives a total resistance (r'_k) of 946 Ω.

In audio, we usually consider frequencies down to 20 Hz (although a 32 foot organ stop will produce 16 Hz) – modern digital sources are certainly capable of sourcing these frequencies, and large transmission-line loudspeakers can reproduce them. In addition, there will be a number of stages to the amplifier, each with filters, so the effect is cumulative. The filter will be made with electrolytic capacitors, which are not known for their initial tolerance or stability of value, so the filter frequency should be much lower than any other filter frequency in the amplifier in case it changes. It has also been argued that a good low frequency response is required not merely to maintain correct amplitude response, but to ensure that the effects on phase and transient response (which extend in-band to a factor of ten times the filter cut-off frequency) are kept to a minimum. Bearing all these factors in mind, it is usual to design for a cut-off frequency of 1 Hz, so:

$$C_k = \frac{1}{2\pi f r_k} = \frac{1}{2 \times 3.14 \times 1 \times 2380} = 170\,\mu F$$

Note that we had to convert the 2.38 kΩ cathode bias resistor back into ohms, and that C is in farads. The nearest standard value to 170 μF is 220 μF, and this is what we would use. You will note that quite a large value of capacitance is needed; valves with a lower r_k are not uncommon, and require a correspondingly larger value of capacitor.

Choice of value of grid leak resistor

Although we have shown the grid leak resistor in place previously, we have not assigned it a value. Historically, it has generally been 1 MΩ for small-signal stages, but somewhat lower for power stages.

It is in our interests to make the grid leak resistor as large as possible, for two reasons.

The grid leak resistor forms a potential divider in conjunction with the output resistance of the preceding stage, and therefore causes a loss of gain. This loss is generally quite small, but it accumulates, so that at the output of a four-stage amplifier, the gain could be significantly less than predicted if this loss is not taken into account.

The second reason is that a large value of grid leak resistor allows the inter stage coupling capacitor to be as small as possible for a given low frequency cut-off.

Once again, if we consult the valve data sheet, we find that there is a limit on the maximum value of grid leak resistor. Usually, two values are given, one for cathode bias, and one for grid bias, the value for grid bias is invariably

lower (2.2 MΩ versus 22 MΩ for the ECC83). The reason that the grid bias value is so much lower is that there is *no* stabilization of operating conditions in this mode. We set grid voltage and the anode current is solely determined by the characteristics of that particular valve.

The clue to the interaction between the valve and the resistor is in the name 'grid leak'. In practice there is always a very small leakage current flowing from the grid to ground, partly because there will always be some contamination of the grid with the oxide coating used to form the cathode emissive surface, but also because of gas current.

Gas current occurs because there is always residual gas in the valve. Brownian motion ensures that individual gas molecules are distributed evenly within the valve, so some must be in the electron path. When a high velocity electron strikes a gas molecule it may have sufficient energy to displace an electron from the molecule's outer shell. The resulting two electrons then continue their path to the anode, but the gas molecule is now positively charged (because it has lost an electron) and is repelled by the anode, so it travels towards the grid/cathode. When the charged molecule, known as an *ion*, strikes the grid, an electron immediately flows up through the grid leak resistor to discharge it. Thus, gas current is also known as ion current.

The flow of discharging electrons is known as leakage current and produces a potential across the grid leak resistor, and the grid becomes positive. V_{gk} is therefore reduced and if the value of grid leak resistor is sufficiently high, this change in V_{gk} becomes significant, and anode current rises. The increase in anode current raises the internal temperature of the valve, releasing yet more gas from the hot structures, further increasing ion current, V_{gk} falls further, the cathode emits more electrons, and the process becomes self-sustaining until the valve is destroyed. Statistically, as more electrons flow from cathode to anode, random collisions between electrons and gas molecules become more likely, so positive ion current into the grid increases with anode current.

However, although grid ion current reduces V_{gk}, which increases anode current, the increased anode current has a counterbalancing effect because it *increases* V_{gk} due to the voltage drop across the cathode bias resistor R_k.

Mullard[1] published a method for determining the maximum permissible value for the grid leak resistor under actual operating conditions. To determine the maximum permissible grid leak resistance of our ECC83 amplifier, we need to know R_k and gm ($R_k = 1.56\,k\Omega$, $gm = 1.54\,mA/V$). First, the effective DC cathode resistance of the circuit is found using:

$$R_{k(effective)} \approx R_k + \frac{1}{gm} = 1.56 + \frac{1}{1.54} = 2.21\,k\Omega$$

Knowing $R_{k(effective)}$ and *gm*, we refer to the graph to find the ratio by which the maximum fixed bias grid leak resistance may be multiplied. See Fig. 2.12.

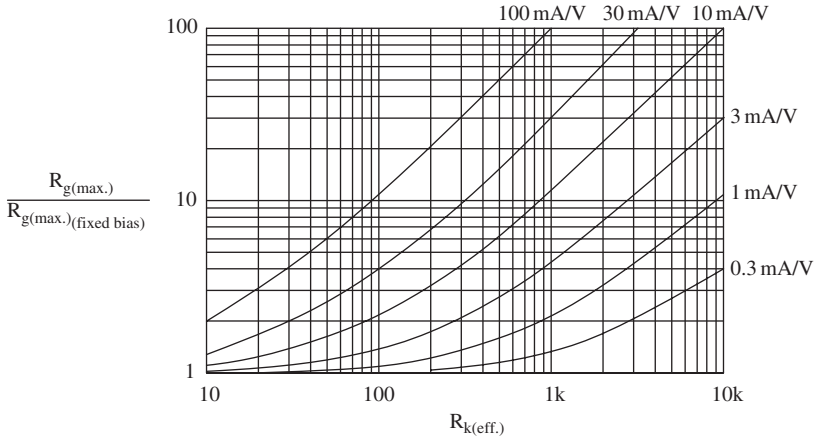


Fig. 2.12 *Maximum value of grid leak resistor (after Mullard)*

Interpolation of the curves suggests that this particular circuit may use a grid leak resistance a factor of four times greater than the maximum fixed-bias grid leak resistance $(2.2\,M\Omega) = 8.8\,M\Omega$.

Nevertheless, the author has seen even larger grid leak resistances in some designs.

Choice of value of output coupling capacitor

This is actually something of a misnomer, since it is actually protecting the input of the next stage from the anode voltage of the first stage, but because the input of a valve stage is usually DC coupled, the coupling capacitor becomes associated with the preceding stage.

The first and most obvious point to observe is that the capacitor should be able to withstand the anode voltage applied to it. What is not so obvious is that it should also be able to withstand the maximum *likely* HT voltage. Modern amplifiers are frequently built using silicon rectifiers for the HT. This means that at the instant of switch-on, the cathodes of the valves may be cold, causing zero anode current. Because the HT is unloaded, it rises to its

maximum possible value, and this voltage appears directly across the coupling capacitors. If they fail catastrophically, then as the valves begin to warm up, the large positive bias on their grids causes them to conduct heavily. The valves may then be destroyed.

Using higher voltage capacitors may be more expensive, but it is cheaper than having to replace an expensive valve (or loudspeakers).

The only other way around this is to ensure that the HT is *never* present before the heaters are warm. Usually this means leaving the heaters on permanently; which is practical, and beneficial, for pre-amplifiers, but we would not wish to leave power amplifier heaters on permanently. A delay is needed, and valve rectifiers were the traditional solution (see Chapter 5).

The other choice is the value of capacitance of the capacitor. Since we will use either poly-something or paper capacitors, which are stable in value, we do not mind if they define the low frequency cut-off of the amplifier. However, all the other arguments used for the cathode decoupling capacitor still apply, and so a choice of 2 Hz for cut-off frequency is not unreasonable. Incidentally, the traditional value of 1 MΩ grid leak and 0.1μF forms a filter whose −3 dB point is 1.6 Hz. Some modern designs use much larger values, and we will consider the rationale for this later.

Miller capacitance

So far, we have looked at the external, wanted, components of our amplifier stage. We will now turn to an unwanted component *Miller capacitance*.

There will always be some capacitance between the anode and the control grid. In a tetrode or pentode it is still there, but greatly reduced. This capacitance is reflected into the grid circuit and forms a low-pass filter in conjunction with the output resistance of the preceding stage. See Fig. 2.13.

We now have two identical stages of the type that we have just designed, connected in *cascade* to form a two-stage amplifier.

Miller capacitance acts like this. When the second valve amplifies the signal, its changing anode voltage is forced to charge and discharge the anode to grid capacitance C_{ag}. That charging current cannot flow into the grid (because the grid is high resistance), so it must be sunk or sourced by the preceding stage. Now, suppose that this capacitance requires a current i to charge it to 1 V. We apply a 1 V step to the input of the amplifier, and the anode moves negatively by 1 V × the gain of the amplifier, in this case by −72 V. The *total* voltage change across the capacitor is therefore $(A + 1)\text{V} = 73\,\text{V}$.

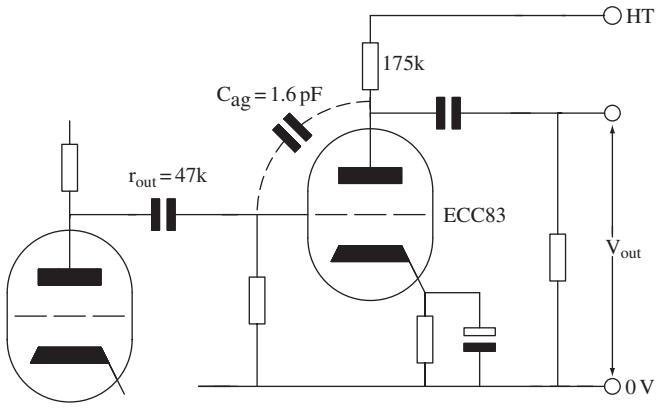


Fig. 2.13 *Miller capacitance*

The total current required to be sunk or sourced by the preceding stage is now $(A + 1) \times i$, or $73i$. We could now reflect this capacitance into the grid by saying that exactly the same current would flow from the source if there was a capacitance between the grid and ground that was $(A + 1)C_{ag}$, hence the Miller equation:

$$C_{Miller} = (A + 1) \cdot C_{ag}$$

It is clear that a quite small value of anode to grid capacitance can have an alarming effect on the high frequency response of an amplifier combination. In our particular case, we find that we have a Miller capacitance of 115 pF (C_{ag} = 1.6 pF for ECC83/12AX7). In combination with the output resistance of the previous stage, this gives a high frequency (HF) −3 dB point of 29 kHz. Stray capacitance will reduce this frequency even further.

There are various ways that we can combat this problem.

- Reduce the output resistance of the preceding stage.
- Reduce significantly C_{ag} by screening the grid support rods from the anode.
- Dramatically reduce C_{ag} by screening the entire grid from the anode (tetrode or pentode).
- Reduce the gain to the offending anode (cascode or cathode follower).

This problem of HF response is so important that we will investigate all of these methods of improving the performance of the common cathode triode stage.

Reducing output resistance of the previous stage

Choosing an E88CC/6922 and operating it correctly reduces the output resistance to a typical value of 10 kΩ. If we also change the second stage to E88CC/6922, the Miller capacitance is then lower, typically 50 pF (due to the gain falling to 30), giving an $f_{-3\,dB}$ point of approximately 300 kHz. However, we have reduced the combined gain of the two-stage amplifier from 5184 (72^2) to 900 (30^2).

As an alternative, we could place a cathode follower (which we will investigate later in this chapter) between the two stages. A cathode follower easily achieves $r_{out} \approx 1\,k\Omega$, so even driving 115 pF of Miller capacitance gives an $f_{-3\,dB}$ point of 1.4 MHz.

Guided-grid, or beam, triodes

In the effort to obtain high μ and *gm* simultaneously, the spacing between anode and grid must be reduced, forcing C_{ag} to rise, and Miller effect causes the stage to have high input capacitance.

Although we might think of the grid as being a mesh of fine wires between the cathode and anode, it has to be supported by rigid vertical metal rods which must be of a much greater diameter than the grid wires in order to determine the cathode/grid spacing precisely. As an example, the grid wire of the 417A triode is specified[2] to be 7.4 μm in diameter, and wound with a pitch of 0.065 mm per turn. On dissection, 80 grid wires were counted under a travelling microscope. The width of the anode near to the grid is ≈3 mm, so the total grid wire area is $80 \times 3 \times 0.0074 = 1.78\,mm^2$. The support rods were measured to be 0.875 mm diameter, the length of the anode near to the rods is 5 mm, and there are two rods, so the total rod area is $2 \times 0.875 \times 5 = 8.75\,mm^2$. The grid support rods have a surface area five times that of the grid wires, and consequently five times the capacitance to a flat anode.

Valves for use at high frequencies seek to minimize capacitance between the anode and support rods, hence the bath-tub anode on the 417A which brings the anode close to the grid wires and avoids the rods, but this still means that a substantial proportion of C_{ag} is due to a structure that has no effect on the passage of electrons, and could therefore be screened from the anode with impunity (See Figs. 2.14 and 2.15).

Logically, these should be called screened-grid valves, but the term had already been used for tetrodes, so the far less satisfactory terms *guided-grid* triode or *beam* triode were used. Valves such as the PC97, PC900 and 6GK5 have internal support rod screens and bath-tub anodes causing C_{ag} to

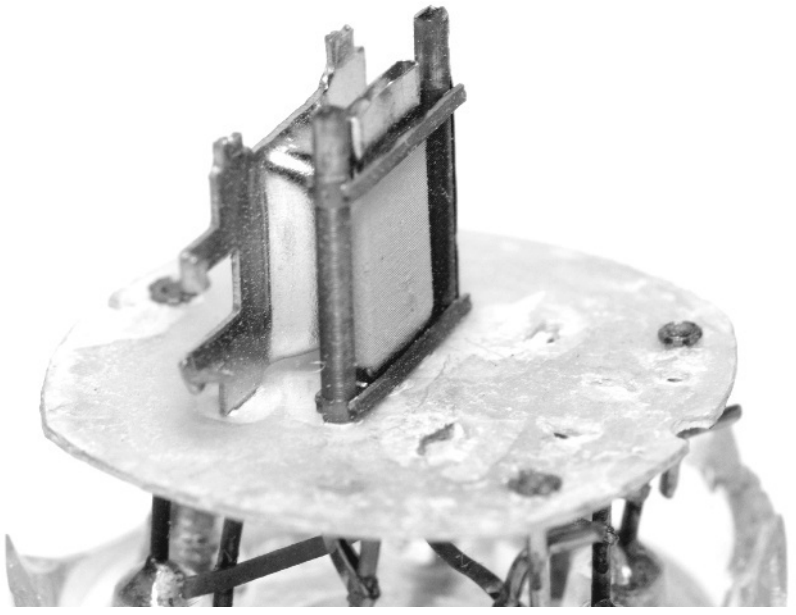


Fig. 2.14 *Dissected 417A: note the size and shape of the remaining anode structure relative to the active area of the grid*

fall to <0.5 pF – a very worthwhile improvement. Sadly, most of these Ultra High Frequency (UHF) valves were designed to be variable μ valves, to allow Automatic Gain Control (AGC), and we will see later that this causes distortion.

The tetrode

As described by the patent, the tetrode[3] was invented to overcome the reduction in gain caused by the electric field of the anode interacting with the electric field of the grid. An auxiliary, or *screen grid*, g_2, is placed between the anode and grid to screen the changing anode potential from the grid. In order to maintain electron flow to the anode, g_2 is connected to a positive potential slightly lower than the anode, so that electrons are attracted to g_2, but most pass through the (coarse) mesh to be captured by the anode as anode current. See Fig. 2.16.

Although originally devised to increase voltage gain, the far more important effect of adding g_2 is that it screens the anode from the grid at AC and

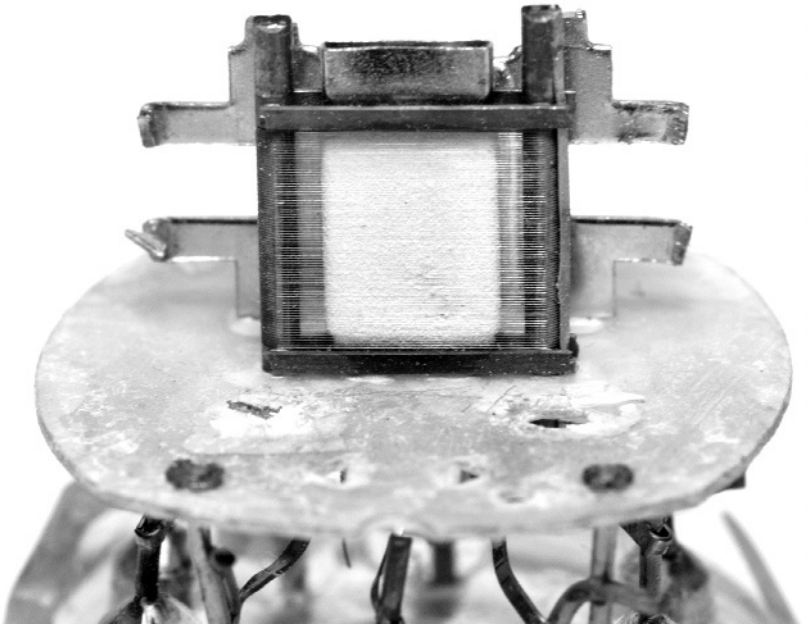


Fig. 2.15 *Dissected 417A: note the relative size of the control grid support structure compared to the grid wires*

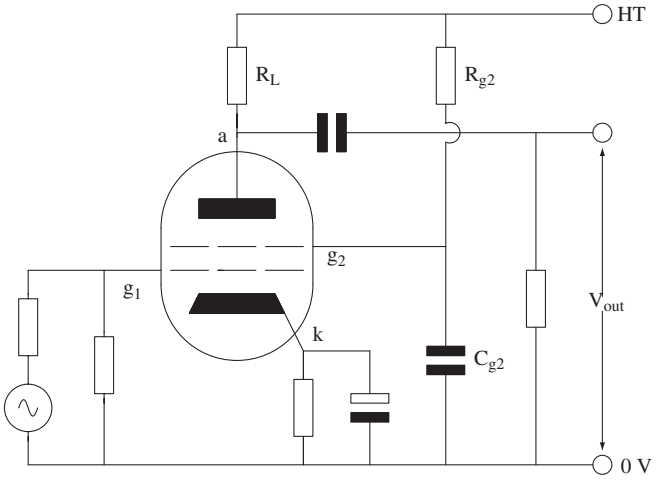


Fig. 2.16 *The tetrode*

greatly reduces Miller capacitance, allowing useful amplification at much higher frequencies. It will come as no surprise to learn that this tinkering with the internal structure to increase gain changes the anode characteristics of the valve. See Fig. 2.17.

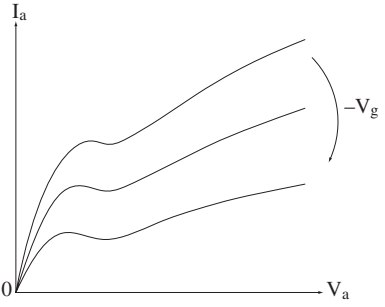


Fig. 2.17 *Anode characteristics of the tetrode*

The kink in the curves is caused by *secondary emission*. At very low anode voltages, electrons are emitted by the cathode in the normal way, and collected by the anode. At slightly higher anode voltages an electron may hit the anode with such velocity that instead of merely being absorbed by the anode, it dislodges *two* low velocity electrons, which are easily attracted to the higher potential of the screen grid. The anode has effectively emitted one electron, and anode current has fallen. As anode voltage rises still further, although electrons are dislodged from the anode, they swiftly return to the anode because the screen grid is at a low potential relative to the anode, so anode current rises once more.

Not only does the kink in the anode curves cause distortion of the signal, but it also implies negative anode resistance, which can cause stability problems, so the pure tetrode was soon superseded.

The beam tetrode and the pentode

These two valves sought to keep the advantages of the tetrode (low C_{ag}) without the disadvantage of the kinked anode characteristic. The pentode works by placing a very coarse grid, the *suppressor grid*, g_3, connected to the cathode, between g_2 and the anode, in order to screen g_2 from the anode. The result of this is that the high velocity electrons emitted from the cathode pass straight through the suppressor grid, but the low velocity secondary electrons

emitted from the anode are screened from g_2, and return to the anode. Because secondary emission electrons from the anode are no longer attracted to g_2, the kink in the anode characteristics of the tetrode is avoided.

Operation of the beam tetrode is different from the pentode in order to avoid infringement of the 1928 Philips pentode patent.[4] Instead of the electrons leaving the cathode from all points of the compass and flowing to the anode, the electrons are directed into two narrow beams of high electron density by the *beam anodes*, which are connected to the cathode. Each beam is further focused and divided into thin horizontal sheets because the g_1 and g_2 windings are vertically aligned, which increases electron density still further. Electrons attempting to leave the anode by secondary emission are now repelled by the incoming flood of electrons and are quickly returned to the anode. Because the dynamics of this space are very similar to the space-charge stabilized emission from the cathode, it is known as a *virtual cathode*. Interestingly, the *pentode* patent hints at a virtual cathode as a means of suppression of secondary emission.

Some electrons may succeed in leaving the anode and travelling a limited distance, and to avoid them reaching g_2, the anode to g_2 distance is rather greater than in the pentode (which is why the anode of the beam tetrode KT66 is larger than that of the pentode EL34, despite their very similar ratings).

The necessary alignment of the g_1 and g_2 windings in a beam tetrode focuses the streams of electrons such that they mostly pass between the wires of g_2, thus reducing g_2 current compared to the pentode, which improves efficiency in a power valve, although there is no reason why a pentode should not adopt the same strategy. In practice, when we come to use a beam tetrode, or a pentode, we see little difference in their electrical characteristics and so we can treat them both as pentodes. (Thorn–AEI classified the PCL82 as a triode/beam tetrode, yet Mullard classified it as a triode/pentode.)

The beam tetrode offers some interesting possibilities. For instance, if we had *two* anodes, and individually connected beam anodes, we could modulate the voltage between the beam anodes to control the ratio of current split between the two anodes. The 6AR8 is such a valve, and was designed mainly for television colour decoding, but its characteristics were also exploited by audio tuners in the mixer stage and in the stereo decoder.

The significance of the pentode curves

When we investigate the anode characteristics of the EF86 small-signal pentode for $V_{g_2} = 100\,\mathrm{V}$, we will see that the anode curves are nearly horizontal. We can draw useful conclusions from this fact.

First, pentode characteristics are very similar to bipolar transistor characteristics, and indicate an anode resistance that is sufficiently high that for most practical purposes, it may be taken to be infinite. The output resistance of the pentode stage is therefore $\approx R_L$.

Second, the anode is able to swing much closer to 0 V than the triode, and so we can obtain a greater peak-to-peak output voltage. This has significant implications for efficiency, and makes the pentode a good choice for high voltage stages.

Third, the shape of the V_a, I_a curves for the pentode (and transistor) is exponential so that:

$$I_a \propto (1 - \exp^{-kV_a})$$

This relationship not only results in the pentode producing significant odd harmonic distortion, but the harmonics extend far further up the spectrum than a triode. As an example, an E55L pentode was biased to $I_a = 50$ mA with a 4k7 anode load from a 410 V supply. The stage clipped at $\approx 73\,V_{RMS}$, so it was tested for distortion at an output of $\approx 50\,V_{RMS}$, whereupon the stage produced 1.3% THD, but note that the distortion spectrum contains significant harmonics up to the 12th. See Fig. 2.18.

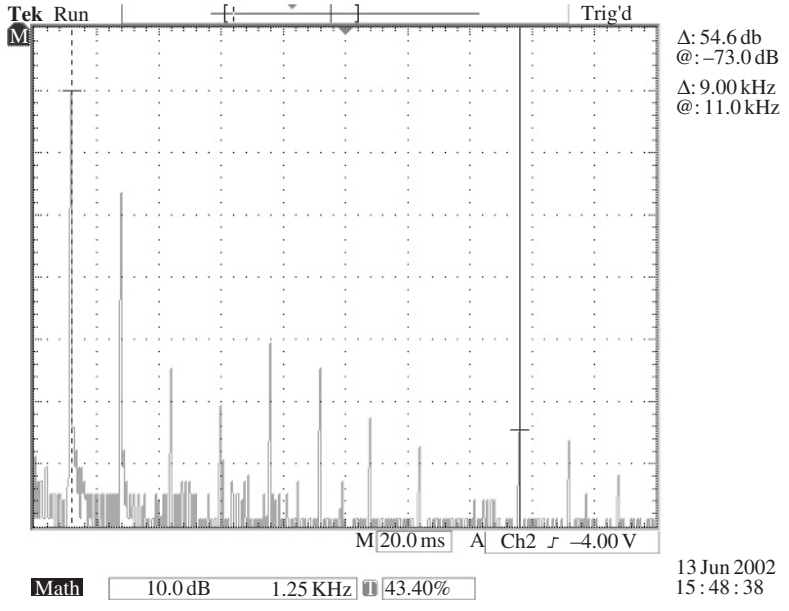


Fig. 2.18 *E55L pentode distortion spectrum*

By contrast, the shape of the triode anode curve is a power law:

$$I_a \propto V_a^{\frac{3}{2}}$$

This equation can be approximated using a binomial series, and although it contains both odd ($x^3, x^5, \ldots$) and even ($x^2, x^4, \ldots$) terms, indicating odd and even harmonics, the terms die away very rapidly (the author has not normally needed to look beyond 6th harmonic when testing triodes). We can therefore expect the triode to produce predominantly 2nd harmonic distortion.

The type of distortion produced is significant because the ear is far more tolerant of even harmonic distortion than odd, partly because the ear itself produces even harmonic distortion, but also because the higher odd harmonics are no longer musically related to the fundamental and sound discordant. The measured distortion of a pentode amplifier must therefore be much lower than the measured distortion of a triode amplifier, because the subjective effect is so much greater, and such amplifiers generally use plenty of negative feedback.

Using the EF86 small-signal pentode

We can now consider how we would use the EF86. R_L is chosen in the normal way, in conjunction with loadlines and the 210 V HT; in this example $R_L = 47\,k\Omega$ and the operating point is at 108 V. See Fig. 2.19.

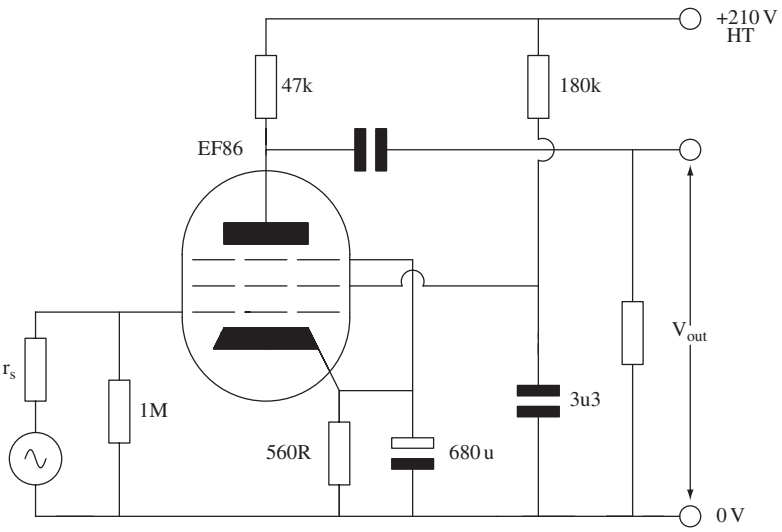


Fig. 2.19 *A small-signal pentode amplifier*

When we come to calculate the gain, we find that the anode characteristic begins to curve as we reach its intersection with the loadline. It is perfectly valid to treat the anode curve as a straight line, and to project this line onto our loadline in order to find the *small-signal* gain, thus giving a gain in this example of 90. See Fig. 2.20.

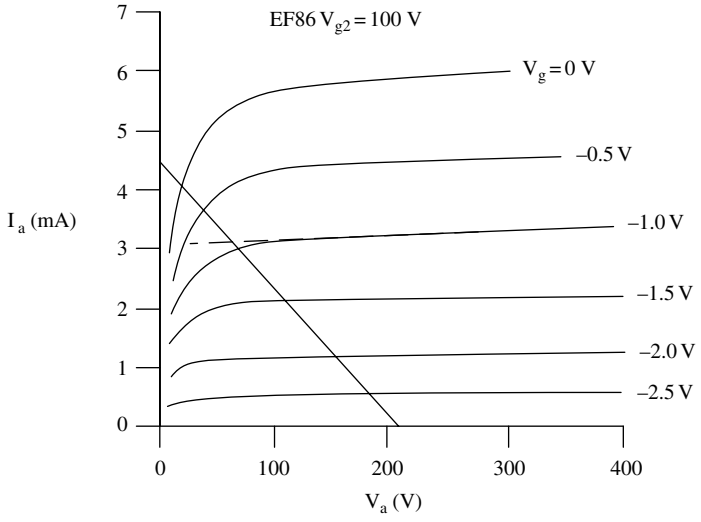


Fig. 2.20 *Anode characteristics and determination of gain of the pentode*

R_{g_2} is chosen either by a detailed perusal of the full data sheets, or by observing that, in general, the anode current is a fixed ratio of the g_2 current. For the EF86 this ratio is $\approx$4: 1. Therefore, if the anode voltage and the g_2 voltage are to be the same, the g_2 resistor should be equal to $4R_L$, and 180 kΩ is therefore appropriate. The latter method is much quicker, but for power valves we must resort to the data sheet.

Although termed a grid, g_2 behaves as an anode in that it receives electrons, and it must therefore have an 'anode' resistance. We need to know this resistance in order to calculate the value of capacitor required to hold g_2 at AC ground potential. Unfortunately, the data sheets for pentodes do not always give $\mu_{g_1-g_2}$, gm_{g_2}, or r_{g_2}, but these can be deduced from triode connected valve data (g_2 connected to anode):

$$\mu_{g_1-g_2} \approx \mu_{triode}$$

Remembering that *gm* describes the controlling effect of V_{gk} on I_k, once the electrons have left the control grid/cathode region, their numbers are fixed, and the density of the g_2 mesh simply determines how the cathode current is split between anode and g_2. Thus:

$$gm_{g_2} = \frac{I_{g_2}}{I_a} \cdot gm_a$$

$$r_{g_2} = \frac{I_a + I_{g_2}}{I_{g_2}} \cdot r_{a(\text{triode})}$$

Using the *triode* curves for the EF86, at $V_a = 108\,V$, $V_g = 1.5\,V$, $r_a \approx 14\,k\Omega$, so $r_{g_2} \approx 70\,k\Omega$. This $70\,k\Omega$ is in parallel with R_{g_2} ($180\,k\Omega$), giving a final resistance of $\approx 50\,k\Omega$, and so for a 1 Hz cut-off, $C_{g_2} = 3.3\,\mu F$.

For a pentode, $I_k \neq I_a$, and we must sum I_a (2.17 mA), and I_{g_2} (0.54 mA), to find I_k (2.71 mA), before we can calculate $R_k \cdot V_{gk} = 1.5\,V$, so the cathode bias resistor must be $560\,\Omega$.

Evaluating *gm* from the anode characteristics, by holding anode voltage constant, and measuring the change in anode current for grid voltage, produces a value of about 1.95 mA/V. For the pentode, the cathode resistance $r_k = 1/gm$, and allowing for the $560\,\Omega$ R_k, we would need a 680 μF decoupling capacitor for a 1 Hz cut-off.

We can also use this value of *gm* in an alternative method of calculating the gain, which is given by the following equation:

$$\begin{aligned} A_v &= gm \cdot R_L \\ &= 1.95 \times 47 \\ &= 92 \end{aligned}$$

The loadline gave a gain of 90, so the agreement is good. Note that this equation does not work for triodes because it assumes infinite r_a.

C_{ag} for the EF86 is given as < 50 mpF, which is a rather quaint way of writing 50 fF (femtofarads, 10^{-15} F). You might wish to consider how Mullard measured a value of capacitance this small in 1955. Clue: you probably wouldn't measure it directly.

$$C_{Miller} = (90 + 1) \times 50 \times 10^{-15} = 4.6\,pF$$

This is a dramatically reduced value compared to the triode, but because it is so small, we will now have to consider stray capacitances that were previously insignificant.

Since the control grid g_1 is near to the cathode, it must have significant capacitance to the cathode, which, since we have bypassed it with a capacitor, is at ground potential. In the data sheet, a value for C_{in} is given, which is the capacitance from the grid to all other electrodes *except* the anode, and is therefore the value of stray capacitance within the valve. For the EF86, C_{in} is 3.8 pF, which gives a total input capacitance (due to the valve) of 8.4 pF. Realistically, we ought to add a few pF for wiring capacitance, so a value of 11.5 pF would be a reasonable total figure.

The ECC83 triode gave a value of 115 pF, so in this respect, the pentode is ten times better. In summary, the pentode has greater gain, greater output voltage swing, and dramatically reduced input capacitance compared to a triode. So why don't we use them all the time?

We have already seen the undesirable distribution of harmonics in pentode distortion, but the real killer for small-signal pentodes is noise.

The EF86 is described as a 'low-noise pentode', and in a very strict sense this is true, because it *is* low noise by pentode standards. By triode standards, it really isn't very good, because of *partition noise*.

This is the *additional* noise, compared to the triode, that is generated by the electron stream splitting to pass either to the anode or to g_2. This additional noise is related to the ratio of anode to screen grid current and to the mutual conductance of the *screen* grid; typically this makes a given pentode 6 dB to 14 dB noisier than that pentode connected as a triode. (Connected as a triode, the EF86 is actually quite a good triode.) Even worse, partition noise has a $1/f$ frequency distribution, which means that its amplitude rises as frequency falls, which has been found to be particularly irritating to the ear.

The cascode

What we would like is a valve, or a *compound device*, that gives the advantages of the pentode with none of its disadvantages – this compound device is known as the *cascode*. See Fig. 2.21.

The cascode bears considerable similarity to the pentode in that there is an arrangement of components (R_1, R_2, C_1) that looks very much like a screen grid bias supply, and indeed this is what it is. The device has a very high r_a, approximately equal to the r_a of the lower valve, multiplied by ($\mu + 1$) of the upper valve.

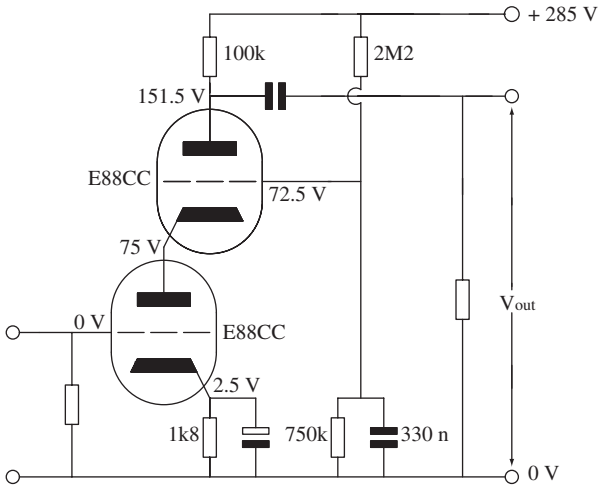


Fig. 2.21 *The cascode*

Operation is as follows: the upper valve has an anode load R_L, as usual, but instead of modulating V_{gk} by varying the grid voltage, and holding the cathode constant, we vary the cathode voltage, but hold the grid constant. The upper grid is biased to whatever voltage we feel is necessary for linear operation of the upper valve, and is held at AC ground by the capacitor. This is significant, because it means that the cathode is screened from the anode by the grid, so Miller capacitance is not a problem. Because we are changing the *cathode* rather than the grid voltage, this part of the stage is non-inverting.

Although the upper valve has a grid in the way of the electron stream, it does not draw current, so partition noise does not occur.

The lower valve operates as a normal common cathode stage, except that it has as its anode load the cathode of the upper valve. Because the dynamic resistance looking into the cathode is low, the gain of the lower valve to its anode is low, so its Miller capacitance is also low. Another way to view the cascode is to consider that both the cascode and the pentode seek to screen the changing voltage across R_L from the sensitive input circuit, and thereby reduce C_{in}. The pentode does this by adding an internal screen between the input grid and anode, and directly reduces C_{ag}, whereas the cascode grounds the grid of the upper valve (which then acts as a screen) and drives the upper cathode from the lower valve.

Because the lower valve has a low value of load resistance, it would generate considerable distortion if it were allowed to swing very many volts. Fortunately, most of the gain is provided by the upper stage, and so distortion of the lower stage should not be a significant problem.

An important point to note with cascodes is that the only general purpose valve that was *designed* to work well in a cascode is the ECC88/6DJ8 or E88CC/6922 (special quality version). Try other valves, by all means, but do not expect the performance to be as good.

We will now see how to design a cascode. It is usual to operate the lower anode at about 75 V, so if we have a 285 V HT, this leaves 210 V across the upper valve. See Fig. 2.21.

We can choose an anode load for the upper valve, and draw a loadline in the usual way. In this case $R_L = 100\,k\Omega$, and $V_g = -2.5\,V$, causing $V_a = -76.5\,V$, which gives a particularly linear operating point. The anode current is therefore 1.34 mA. See Fig. 2.22.

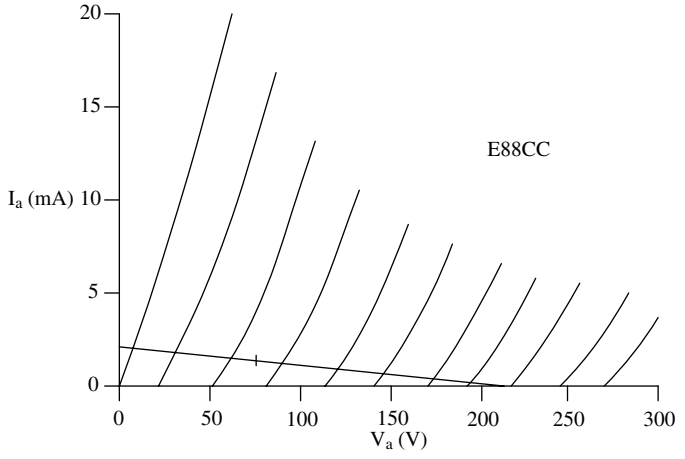


Fig. 2.22 *Choice of operating point of the upper valve of a cascode*

If the anode of the lower valve is to be operated at 75 V, and the upper valve has a V_{gk} of $-2.5\,V$, then the grid of the upper valve must be at 72.5 V. Since the grid of the upper valve does not draw any current, its voltage is set by the potential divider, and completely determines the conditions of the upper stage, which is working in grid or *fixed* bias mode. We still have to be careful not to exceed the maximum permissible grid leak resistance of the upper valve, which for an E88CC/6922 is 1 MΩ, but the Thévenin resistance of the potential divider is 560 kΩ, so we are well within limits. (We have assumed that the DC resistance of the power supply is zero in making this calculation.) We only need a 0.33 μF capacitor to make the grid a short circuit to ground as far as AC is concerned ($f_{-3\,dB} = 1\,Hz$), compared to 3.3 μF for the EF86 g_2 capacitor.

Attempting to investigate the lower stage using anode characteristic curves is not very helpful. Instead, we will use the *mutual characteristics* of anode current against *grid* voltage. See Fig. 2.23.

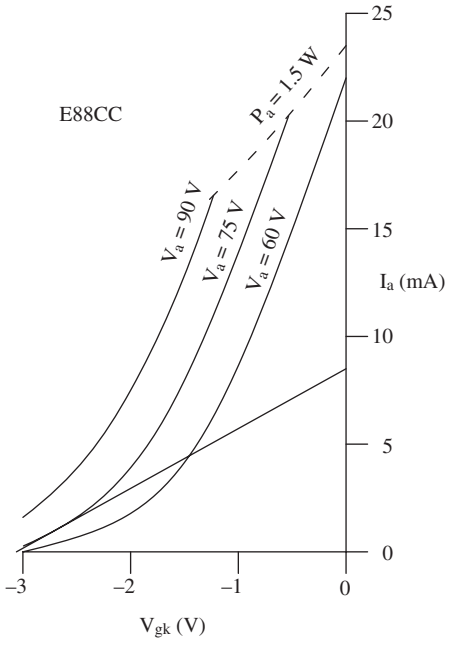


Fig. 2.23 *Triode mutual characteristics*

We know that V_a of the lower valve is 75 V, so we can look along the curve for $V_a = 75$ V until we come to the point where $I_a = 1.34$ mA (upper and lower anode currents are equal), this is the operating point of the lower valve, and gives a V_{gk} of about 2.6 V. Plotting the point $V_a = 75$ V, $I_a = 1.34$ mA on the anode characteristics gives $V_{gk} = 2.4$ V, so the agreement is not too bad. From this we could take an average value of 2.5 V, and calculate the value of R_k at 1.8 kΩ.

Because the cascode is made up of one stage that is non-inverting, and one that inverts, the output is inverted with respect to the input. The gain of a cascode, where V_1 is the lower valve, and V_2 is the upper valve, with equal anode currents is:

$$A_v = \frac{1}{\frac{1}{gm_1 \cdot R_L} + \frac{r_{a2}+R_L}{R_L} \cdot \frac{1}{\mu_1(\mu_2+1)}}$$

This unwieldy equation is frequently approximated to $gm_1 \cdot R_L$. From the equation, we see that we need to find *gm* for the lower valve. This is easily done using the mutual characteristics, by measuring the gradient at the operating point.

$$gm = \frac{\Delta I_a}{\Delta V_g} = \frac{8.35}{3.08} = 2.7\,\text{mA/V}$$

We need r_a for the upper valve, but we are not sitting conveniently on a grid line, so we must interpolate (guess). We can either do this by taking an average of the values either side of the operating point (if they are symmetrically about the operating point), or we could use a French curve to draw a new grid curve where we need it (quite a good method). In this instance, we will take an average value. See Fig. 2.24.

$$r_{a(V_g=-2\,V)} = 5.5\,\text{k}\Omega$$

$$r_{a(V_g=-3\,V)} = 6.45\,\text{k}\Omega$$

Therefore we will say that at $V_g = 2.5\,V$, $r_a \approx 6\,\text{k}\Omega$.

At the operating point of both valves, $\mu = 32.5$.

Putting all these values into the equation yields a gain of 214. Using $gm_1 \cdot R_L$ would have given a gain of 270, which is 2 dB high; nevertheless, the

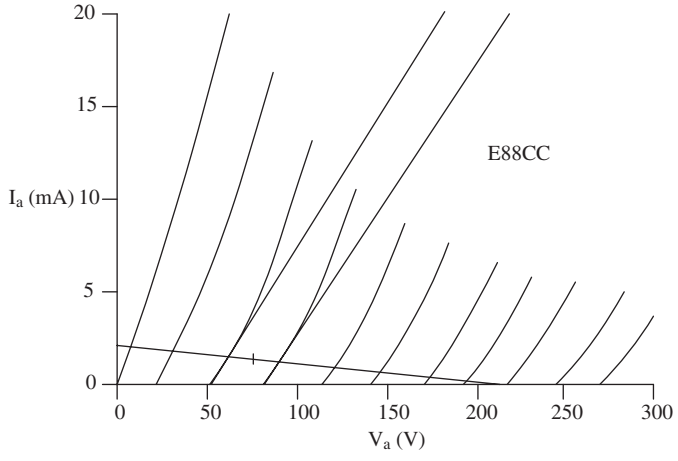


Fig. 2.24 *Averaging two values of r_a to find intermediate value*

approximation is useful, because it tells you whether it is worth pursuing the design any further.

We can now use this value of total gain to calculate the gain of the lower stage (if we wish). This is a useful exercise, because it allows us to find the voltage swing on the lower anode. From this we can check linearity (which might be problematic), and Miller capacitance. We can read the gain of the upper stage from the loadline, which gives us a gain of 30, so the gain of the lower stage must be 7.1. $C_{ag} = 1.4\,pF$ for E88CC, so the Miller capacitance is:

$$C_{Miller} = (7.1 + 1) \times 1.4\,pF = 11.3\,pF$$

As with the pentode, we should add the strays, 3.3 pF for the internal (valve) strays and 3 pF for external strays, to give a total value of 18 pF. This is not quite as good as the pentode that we saw earlier, but if the pentode had used a 100 kΩ anode load, its gain and Miller capacitance would have doubled and the answers would then have been comparable.

The values of cathode bias resistor and decoupling capacitor are calculated in the normal way for a triode.

We need not use equal values of anode current in the upper and lower valves. Adding a resistor from the HT to the lower anode allows additional current to flow into the lower valve. This is useful, because increased lower anode current increases gain (by increasing gm_1) and improves linearity. See Fig. 2.25.

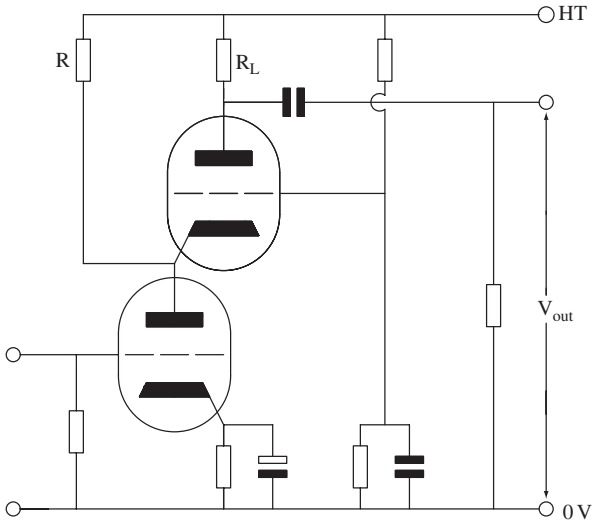


Fig. 2.25 *Increasing I_a of the lower valve in a cascode*

As an extreme example, we might need a low noise low distortion cascode using half of a 6SN7 dual triode as the upper valve, so we might set its anode current to 8 mA (good linearity at this current). However, if the lower valve was a triode-strapped E810F, passing 45 mA, an additional 37 mA would be required. If $V_a = 100\,V$ for the E810F, and the HT = 400 V, then:

$$R = \frac{V}{I} = \frac{400 - 100}{37} = 8.1\,k\Omega$$

All topologies that involve operating cathodes at voltages significantly above ground have problems because of heater/cathode leakage currents and the maximum allowable heater to cathode voltage V_{hk} (see Chapter 4). It is not uncommon for the cathode of a valve to be unbypassed and therefore have signal voltages on it. If, as in the cascode, the gain to the cathode of the upper valve is low, and we are using the device because of its good noise performance, then it is likely that the signal voltage on that cathode is very small, perhaps only a few millivolts. Leakage currents via the heater/cathode insulation become worse as V_{hk} rises, so the combination of $V_{hk} = 75\,V$, and a small signal voltage, means that the effects can be significant. The author once made a circuit using valves that were rated at $V_{hk(max.)} = 150\,V$, operated the valves at $V_{hk} = 120\,V$ and suffered low frequency noise, which was only cured by sitting the relevant heaters on a 150 V DC supply. There is an understandable reluctance to do this, because it means that we need two or more heater supplies, one connected to ground as normal, and another connected to an elevated voltage. We will return to this practical problem later.

The cathode follower

The circuits that we have considered up until now have been concerned exclusively with providing voltage gain. Sometimes we need a *buffer* stage that provides high input and low output resistance. The cathode follower[5] has a voltage gain of slightly less than 1, a low output resistance, typically $\leq 1\,k\Omega$, a high input resistance ($\approx 500\,M\Omega$ in valve microphones), and is non-inverting. We will consider the fixed bias version of the cathode follower first. See Fig. 2.26.

We have changed the position of the load resistor, so that the output is now taken from the cathode, but the circuit can still be analysed in the same way as before, using loadlines. See Fig. 2.27.

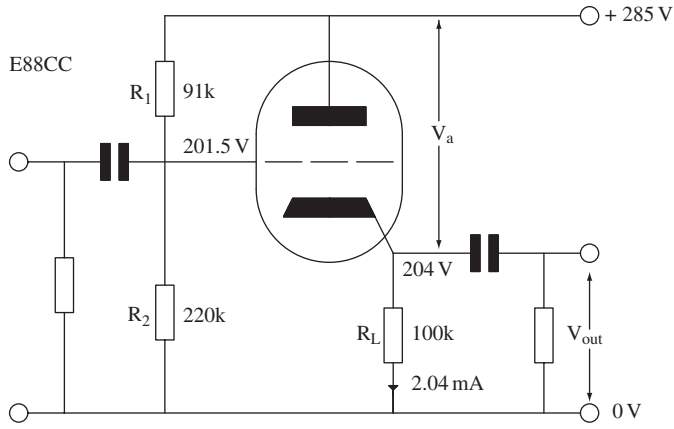


Fig. 2.26 *Fixed bias cathode follower*

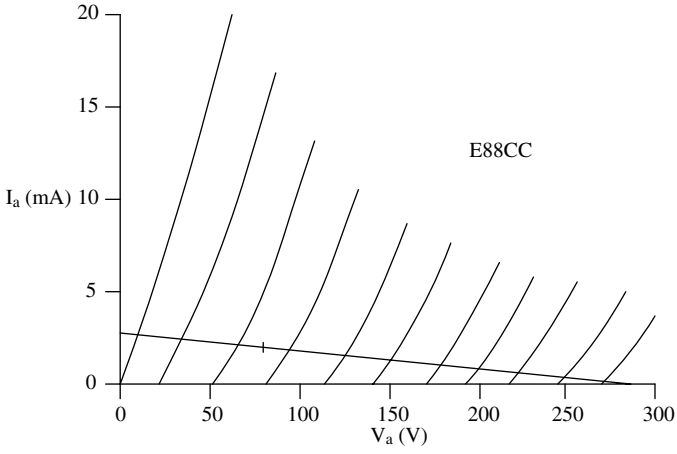


Fig. 2.27 *Operating point of fixed bias cathode follower*

$R_L = 100\,k\Omega$, and so we draw the appropriate loadline, $V_g = -2.5\,V$, with $V_a = -81\,V$, because of the excellent linearity in this region. Remembering that V_a is actually the *anode to cathode* voltage, the cathode is now at $285\,V - 81\,V = 204\,V$, and because $V_{gk} = -2.5\,V$, the grid must be at $201.5\,V$ to bias the valve to this condition. This voltage is set by the potential divider R_1, R_2.

The cathode follower is simply a special case of the common cathode amplifier with 100% negative feedback (parallel derived, series applied). To

find the gain after feedback, we use our normal technique of measuring the gain from the loadline ($A_v = 28.5$), and apply the feedback equation:

$$A_{fbk} = \frac{A_0}{1 + \beta \cdot A_0}$$

Since we have 100% feedback, $\beta = 1$, and the gain of our example becomes $28.5/29.5 = 0.97$.

We saw earlier that the AC resistance at the cathode was:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

But for a cathode follower, R_L from the anode to the HT $= 0$, so this equation can be approximated to $1/gm$. From the anode characteristics, $gm \approx 5$ mA/V, this gives an output resistance of $\approx 200\ \Omega$. This is not a particularly accurate answer, since the method of determining *gm* was crude, but this does not matter, since it is usual to operate an audio cathode follower with $a \approx 1\ k\Omega$ resistor in series with its output to ensure stability – this then swamps the slight inaccuracy. Nevertheless, $1.2\ k\Omega$ is a low output resistance for a valve stage.

As shown, the stage does not have a high input resistance, although this configuration is useful for making Sallen & Key active filters (see Appendix). We need to rearrange our bias to achieve a high input resistance. See Fig. 2.28.

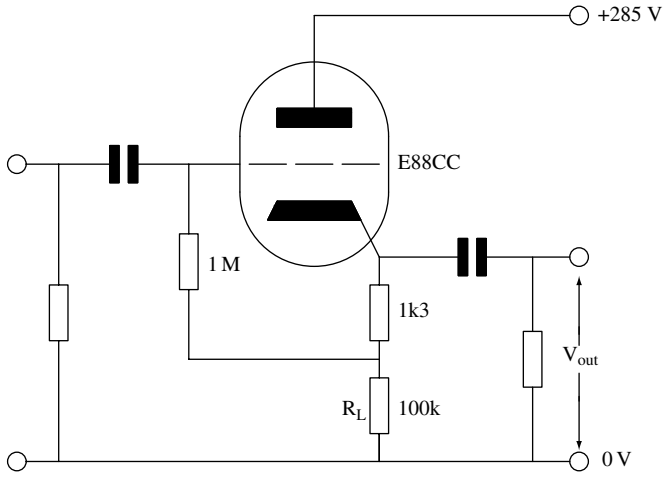


Fig. 2.28 *Cathode bias cathode follower*

We now have cathode or *self* bias provided by the 1.3 kΩ resistor, whose value is calculated in the normal way. You will note that by adding this resistor, we have slightly increased the value of R_L, and indeed this was also the case in the common cathode amplifier, but this ≈1% increase has a negligible effect on circuit conditions.

At first sight, this configuration is very little better than the fixed bias configuration, as the input resistance appears to be only 1.1 MΩ. However, the 1 MΩ grid leak resistor has been *bootstrapped*,[6] which is to say that the entire input signal does *not* appear across it.

It works like this. We have just calculated the gain A_v to the cathode as being 0.97. We can calculate the attenuation of the potential divider formed by the cathode bias resistor and R_L as being 0.987, therefore the proportion of input signal voltage at the lower end of the grid leak resistor is $0.96\,V_{in}$. Now, since the output of a cathode follower is *non-inverting*, this means that there is only $0.04\,V_{in}$ *across* the grid leak resistor. The *signal* current through this resistor will therefore be only 4% of what it would have been, had the grid leak resistor been connected directly to ground. It presents an input resistance equivalent to 1 MΩ/0.04 = 25 MΩ. Formalizing this argument:

$$r_{input} = \frac{R_g}{1 - A \cdot \frac{R_L}{R_L + R_k}}$$

Note that A is the gain of the cathode follower, not the original loadline gain. A similar argument can be used to determine the input capacitance of the cathode follower:

$$C_{input} \approx C_{ag} + (1 - A) \cdot C_{g-k}$$

Note that this is an approximate value because there will be significant strays. Using our example with the E88CC:

$$\approx 1.4\,pF + (1 - 0.96) \times 3.3\,pF = 1.5\,pF$$

We should add a few pF for wiring strays, as we did before, which brings the likely input capacitance of the cathode follower to 4.5 pF, which is rather less than half the value of the cascode or pentode.

It has been suggested that the linearity of the cathode follower is questionable. It is hard to see how this accusation can be true, particularly if the operating point has been chosen carefully, as in the previous example, since the stage operates under 100% negative feedback. This means that any non-linearity will

be reduced in proportion to the feedback factor $(1 + \beta A_0)$, which in our example gives a reduction of 30:1.

Nevertheless, it *is* possible to do even better. We mentioned earlier that μ was one of the more stable valve parameters, whereas r_a varies considerably with anode current. This is significant, because it is mostly the variation of r_a that causes distortion, and we can see why this is if we look at the equation for the gain of a common cathode amplifier:

$$A_v = \mu \cdot \frac{R_L}{R_L + r_a}$$

If we could make R_L very large, ideally infinite, r_a would become insignificant by comparison and could no longer cause distortion. Provided that we have chosen a sensible operating point where μ does not vary greatly, we will then have a very low distortion buffer. Unfortunately, if we simply make R_L very large, we find that there is such a voltage drop across it that we need an HT of over 2 kV! See Fig. 2.29.

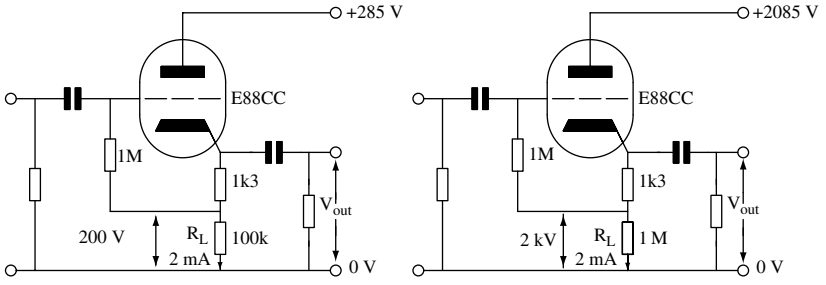


Fig. 2.29 *Effect of increasing R_L in a cathode follower*

We need a way around the problem of excessively large values of R_L, and to do this, we need to examine some definitions.

Sources and sinks: definitions

A current or voltage source is a supply of energy (such as a battery) capable of supplying energy into a load whose other terminal is connected to ground, whereas a sink may *control* the characteristics of an external source of energy, but provides none of its own. Audio electronics often needs real world approximations to these hypothetical devices in order to improve the AC

performance of the surrounding circuit, so the following definitions are couched in terms of their common effects on AC performance, even though practical implementations using active devices may be equally effective at DC.

A perfect constant voltage source/sink is a short circuit (zero resistance) to AC, and Ohm's law therefore ensures that even an infinite AC current passing through it causes zero AC voltage across it. Although active devices such as voltage regulators are frequently better, suitably sized capacitors are often inescapably used as AC approximations to constant voltage sources/sinks. Common audio examples include the reservoir capacitor in a capacitor input power supply (source), or the cathode bypass capacitor (sink).

Conversely, a perfect constant current source/sink is an open circuit (infinite resistance) to AC, and even an infinite AC voltage across it is incapable of causing an AC current to pass through it. Active constant current sources/sinks are becoming more common, but inductors are also used as AC approximations to constant current sources/sinks, the main audio examples being the choke in a choke input power supply (source), or the primary inductance of an output transformer (sink).

Although it has been suggested that capacitors and inductors may be used as approximations to perfect sources or sinks, the implication is that topologies traditionally using these approximations may be replaced by active devices which can be more nearly perfect. It is probably true to say that the primary difference between a valve amplifier designed in the 'Golden Age', and a modern design is that the modern design is likely to replace passive components with active devices to approximate perfect sources and sinks more closely.

The common cathode amplifier as a constant current sink

We saw earlier that leaving a common cathode amplifier with R_k unbypassed caused r_a to rise due to negative feedback. We can exploit this effect deliberately to create a constant current sink. See Fig. 2.30.

Let us suppose that we need to sink a current of 2 mA, using an E88CC valve and that we have 204 V of HT available for the sink. We can treat $V_a = 204\,V$, $I_a = 0\,mA$ as one end of a loadline, and plot this point on the graph. See Fig. 2.31.

Plotting $V_a = 204\,V$, $I_a = 0\,mA$ is easy, but we don't yet know where the other end of the loadline will be. However, we *do* know a point on the

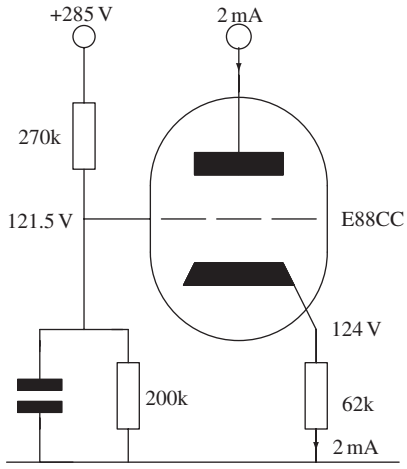


Fig. 2.30 *Constant current sink*

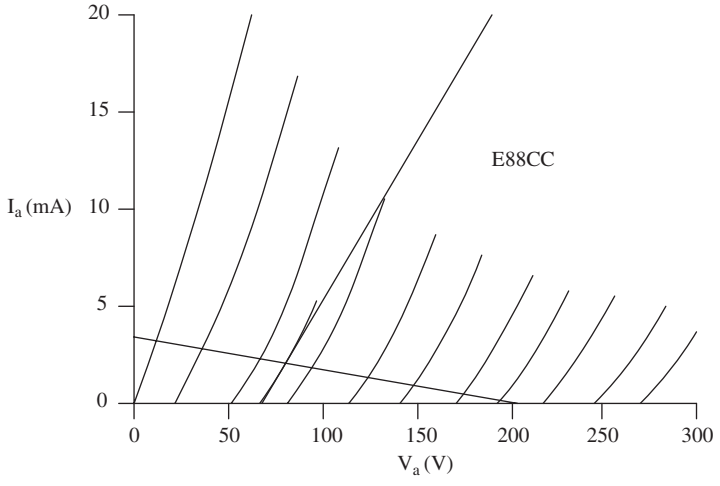


Fig. 2.31 *Operating conditions of constant current sink*

line; we know that $I_a = 2$ mA at the operating point, although we do not know the voltage. It is up to us to *choose* a voltage, and $V_a = 81$ V is a good choice for linearity. Linearity is still important in a constant current sink because the complete circuit will probably modulate the anode voltage with an audio signal. If linearity is poor, this indicates non-constant r_a, which is part of the term that governs the output resistance of the sink. If the output resistance varies with applied voltage whilst it is being used as an active load for another

valve, it will cause distortion in that valve. If we now draw our loadline, we can find the current through R_L when $V_a = 0$. From this we can calculate the value of R_L, which is then 60 kΩ. The nearest value to this is 62 kΩ, and this is what we would use.

Since $I_a = 2$ mA, we know that the cathode of the valve will be at 124 V. $V_{gk} = -2.5$ V, so the grid needs to be at 121.5 V. This voltage is set in the normal way using the potential divider and capacitor combination.

The AC resistance, looking into the anode of this circuit, is:

$$r_{sink} = r_a + (\mu + 1) \cdot R_k$$

For our design, this gives a value of slightly more than 2 MΩ. Achieving this result with a pure resistance would require a 4 kV HT supply. The AC resistance is in parallel with C_{out}, and C_{ag} causes the sink gain to fall, so the sink impedance falls as frequency rises. (C_{out} is the capacitance from the anode to all other electrodes *except* the grid.)

Pentode constant current sinks

Pentodes are even better as constant current sinks because of their high μ, and are particularly useful if the allowable voltage drop across the sink is quite low.

If we needed a constant current sink of 10 mA, but were only allowed a 100 V drop across it, an E88CC could only achieve an output resistance of ≈100 kΩ, which is still a ten-fold improvement over a 10 kΩ resistor, but a pentode can do rather better.

If we leave even a 2 kΩ cathode resistor unbypassed, a pentode can increase its output resistance to >10 MΩ. This is a stunningly good constant current sink, but it should be remembered that pentodes tend to be noisy, so this would not be a good choice in the first stage of a sensitive pre-amplifier. See Fig. 2.32.

When using a pentode as a constant current sink, it is vital to remember that the cathode resistor passes not only the desired constant current, but *also* the g_2 current. Note also that the g_2 decoupling capacitor must be taken to the cathode, and not to ground. This is because we want cathode feedback to increase r_a, but we do not want the voltage between g_2 and the cathode to vary, as this would cause positive feedback that would reduce r_a.

Some pentodes make better constant current sinks than others because their anode characteristics are flatter, giving a higher output resistance, or because the flat part of their anode characteristics swings closer to 0 V. The

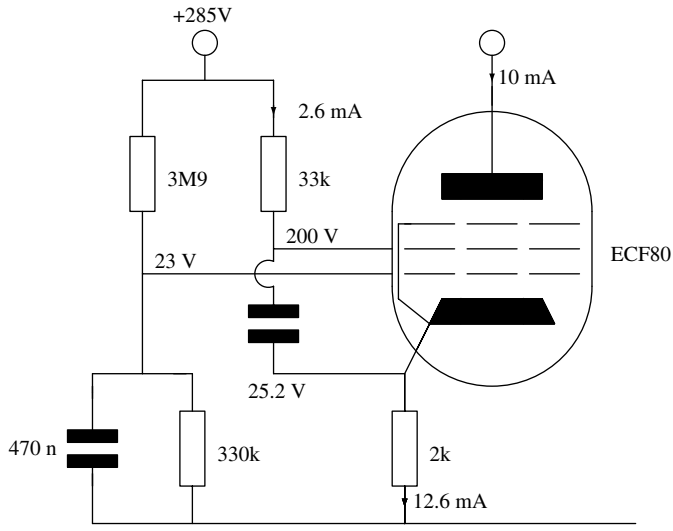


Fig. 2.32 *Pentode as constant current sink*

table shows single pentodes that are particularly suitable in constant current sinks.

Type	Optimum current	C_{out}	P_a
EF91/6AM6	$\leq$6 mA	3.1 pF	2.5 W
EF184/6EJ7	8 mA–15 mA	3 pF	2.5 W
EL83/6CK6	15 mA–30 mA	6.6 pF	9 W
EL822	20 mA–45 mA	6 pF	12 W

The table shows optimum currents much lower than $I_{a(max.)}$, partly because at higher currents, the anode curves tilt away from the horizontal, indicating reduced R_a, but mostly because the major contribution to output resistance actually comes from the unbypassed R_k, whose value is multiplied by a factor of $gm_1 \cdot r_a(\mu)$. Higher currents require less bias and a reduced value of R_k, thus reducing output resistance. For maximum output resistance, it is better to use a valve with an oversized P_a, requiring a large R_k, than a perfectly rated P_a, requiring a smaller R_k. Unfortunately, the disadvantage of a constant current sink operated at very low I_a is that I_{g_2} becomes a significant proportion of I_k, making the circuit inefficient.

As an example, we might require an 8 mA constant current sink. However, this is the lower end of efficient operation for an EF184, and the curves show that it is likely to require $I_{g2} \approx 3$ mA, which means that total HT current has been increased by ≈38%. If there is only one sink, then this isn't a problem, but if there are many such sinks, this can greatly increase the cost of the HT supply. If we were willing to reduce the sink current to 6 mA, this would bring it within range of the EF91, which only requires $I_{g2} = 1.55$ mA in this instance, reducing HT current from 11 mA to 7.55 mA. Although the EF91 does not have quite such attractive curves as the EF184, it is far cheaper to use, and if HT current is limited, it can be worth bending a design to allow its use.

Optimally biased, most small-signal pentodes split I_k between I_a and I_{g2} by ≈4:1 ratio. Thus, 8 mA of I_a typically requires 2 mA of I_{g2}. It is very important to check that the pentode's I_a, P_a, and especially P_{g2} are not exceeded. Successful design of pentode constant current sinks requires full data sheets with curves, or a valve tester/rig to allow voltages to be imposed and currents determined experimentally (which is more reliable).

If the constant current sink is to be used in a stage with a low signal level, it is worth considering hum and screening. The EF184 has an integral metal screen, the EF91 has a conductive paint screen on the inside of the envelope, but the EL83 and EL822 power valves are completely unscreened.

The cathode follower with active load

You will probably have realized that the requirements for the triode constant current sink were set by the cathode follower designed earlier, so we can now combine them to form a cathode follower with an *active load*. See Fig. 2.33.

Because the value of R_L for the upper stage is now so large, the gain becomes:

$$A_v = \frac{\mu}{\mu + 1}$$

The gain is therefore 0.97, which is only a little higher than before, but the distortion is greatly reduced. It is possible to make distortion predictions, but these are of very doubtful value, since *real* valves do not behave in the nice mathematical fashion that equations require to generate sensible answers.

Grid current can cause a cathode follower to have very much higher distortion than expected. The author tested a self-biased cathode follower using a 6C45П with an EF184 constant current sink active load, then determined its

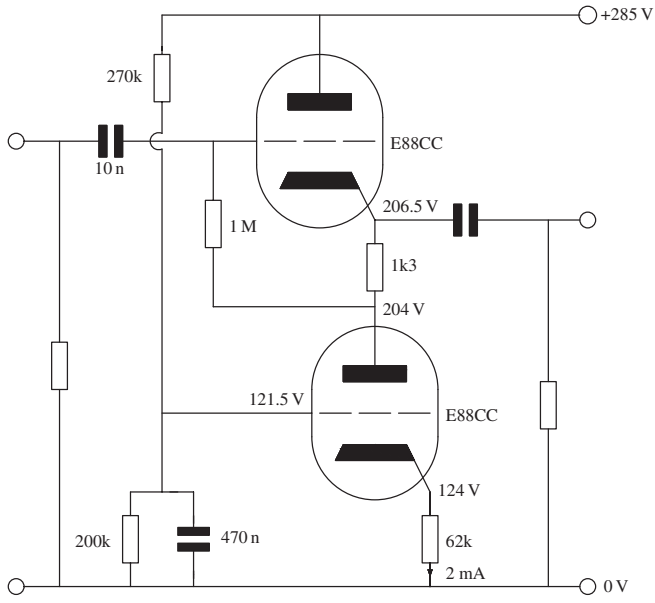


Fig. 2.33 *Cathode follower with active load*

input resistance by measuring the relative loss when driven from a 1 MΩ source resistance compared to 5 Ω. Sadly, the input resistance was not quite as high as predicted, and adjusting the value of grid leak resistor from 150 kΩ to 1 MΩ not only changed the input resistance and slightly changed I_a (indicating grid current), but also reduced the distortion at +20 dBu from 0.23% to 0.052%. Reducing the source resistance from 1 MΩ to 24 kΩ further reduced the THD + N from 0.052% to 0.02%. Cathode followers are often used as buffers after volume controls, so this sensitivity to source resistance can be significant, particularly since we will see in Chapter 7 that some volume controls have higher output resistance than others.

To sum up, a carefully designed cathode follower with a resistive load produces low distortion – replacing this with an active load improves it further, and challenges test equipment, but for optimum distortion the valve should be selected/tested for low grid current.

The White cathode follower

Named after its inventor, the White cathode follower[7] is the basis of all output transformer-less power amplifiers because of its low output resistance.

The circuit comes in two forms, one self-contained, and the other requiring an external phase splitter.

Analysis of the self-contained White cathode follower

The lower valve is fed with a signal from the upper valve, which in turn, it feeds back into the cathode/grid circuit of the upper valve. At the input to the lower valve, the circuit may be considered to be a cascode amplifier. See Fig. 2.34.

Provided that μ is reasonably large, and the cathode is bypassed:

$$A_v \approx gm \cdot R$$

This gain will be devoted to reducing the output resistance at the cathode of the upper valve:

$$r_k = \frac{R + r_a}{\mu + 1}$$

Combining these:

$$r'_k \approx \frac{R + r_a}{(\mu + 1)gm \cdot R}$$

μ is usually rather greater than 1, even for power triodes, and if we substitute $\mu = gm \cdot r_a$:

$$r'_k \approx \frac{R + r_a}{gm^2 \cdot R \cdot r_a} \quad \text{or} \quad \frac{1}{gm^2} \cdot \frac{R + r_a}{R \cdot r_a}$$

We can now recognize the R, r_a term as the inverted parallel combination of R and r_a. This is significant because it indicates that there is a point beyond which increasing R has no effect, and that final output resistance is limited by r_a:

$$r_{out} = r'_k \approx \frac{1}{gm^2 \cdot R'}$$

where:

$$R' = \frac{R \cdot r_a}{R + r_a}$$

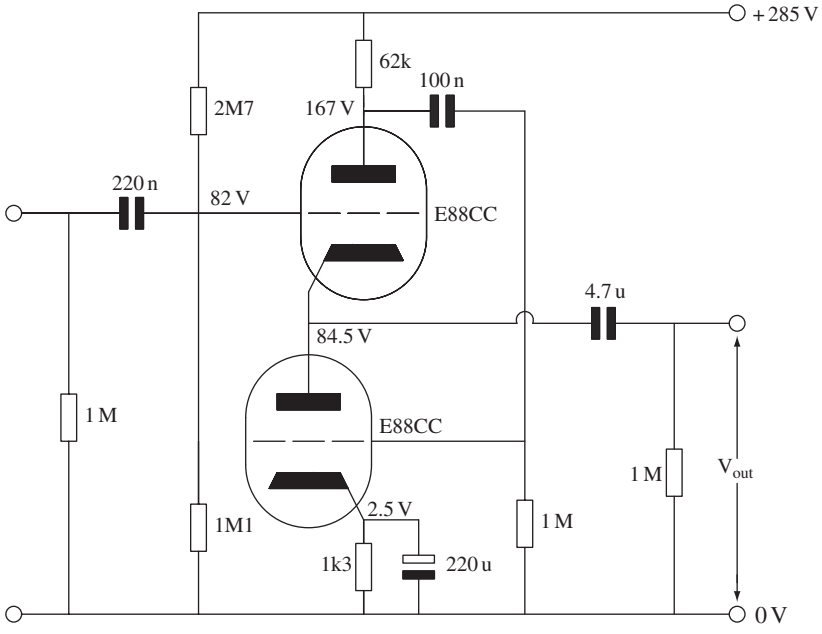


Fig. 2.34 *Self-contained White cathode follower*

It should be noted that two rather dubious approximations were made to derive this result, both of which relied on high μ. The example in Fig. 2.34 was optimized for low output resistance, and $R \approx 10r_a$, beyond which limit, no practical improvement is possible.

Although superficially completely different, the self-contained White cathode follower and the SRPP described later in this chapter are both shunt-regulated amplifiers because the two valves each contribute to the AC load current. Rigorous equations for gain and output resistance of the White cathode follower derived by Amos and Birkinshaw[8] are as follows:

$$A_v = \frac{\mu_1(\mu_2 R + r_{a_2})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

$$r_{out} = \frac{r_{a_2}(R + r_{a_1})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

where V_1 is the upper (amplifying) valve, and V_2 is the lower (regulating) valve.

Using an E88CC as an example, with $gm \approx 5\,\text{mA/V}$ and $\mu \approx 32$, the approximate equation predicts $r_{out} \approx 6.9\,\Omega$, whilst the rigorous equation predicts $r_{out} = 6.7\,\Omega$. Experimentation with a spreadsheet reveals that this version of the White cathode follower is unsuited to low μ valves, since a 6080 ($\mu = 2$) predicts $r_{out} \approx 35\,\Omega$, which is worse than when used as a standard cathode follower ($r_{out} \approx 15\,\Omega$). However, a triode strapped E55L ($\mu = 30$) predicts $r_{out} < 2\,\Omega$, and a triode strapped D3a ($\mu = 80$) can achieve $r_{out} < 1\,\Omega$. It is easy to become excited about low predicted output resistances, but we should always remember that all such equations contain the implicit assumption that the output resistance of the HT supply is $0\,\Omega$, which is normally only achieved by a regulated supply.

Since the White patent suggests that the circuit is particularly suitable for driving analogue video cables (which are typically $75\,\Omega$ transmission lines), it is not surprising that the stage makes an excellent output cable driver for a pre-amplifier.

Note that because the feedback that causes the low output resistance is AC coupled, output resistance rises at low frequencies not to $1/gm$, but to:

$$r_{out(LF)} = \frac{R + r_a}{\mu + 1} \| r_a = \frac{r_a(R + r_a)}{r_a(\mu + 1) + R}$$

In this instance, r_{out} rises to $1.5\,\text{k}\Omega$, rather than $200\,\Omega$, which is what a normal cathode follower would achieve. The practical implication is that the stage will not short circuit induced output cable noise (such as mains hum) as effectively as a stage with a true $6\,\Omega$ output resistance from DC to light.

Usually, we do not need to calculate the gain A_v precisely, and the general cathode follower approximation of $A_v = \mu/(u + 1)$ is perfectly adequate, but if the amplifier was to be used as the basis of a Sallen & Key filter, the rigorous calculation of gain could be needed.

The White cathode follower as an output stage

The primary use of the White cathode follower is as an output stage for output transformer-less amplifiers. A series resistor in either of the HT rails is a serious waste of power, so we must use the version preceded by a phase splitter. See Fig. 2.35.

Assuming that neither valve switches off under any signal conditions (Class A), the gain of the lower valve is:

$$A_{v(lower)} = \frac{\mu \cdot R_L}{R_L + r_a}$$

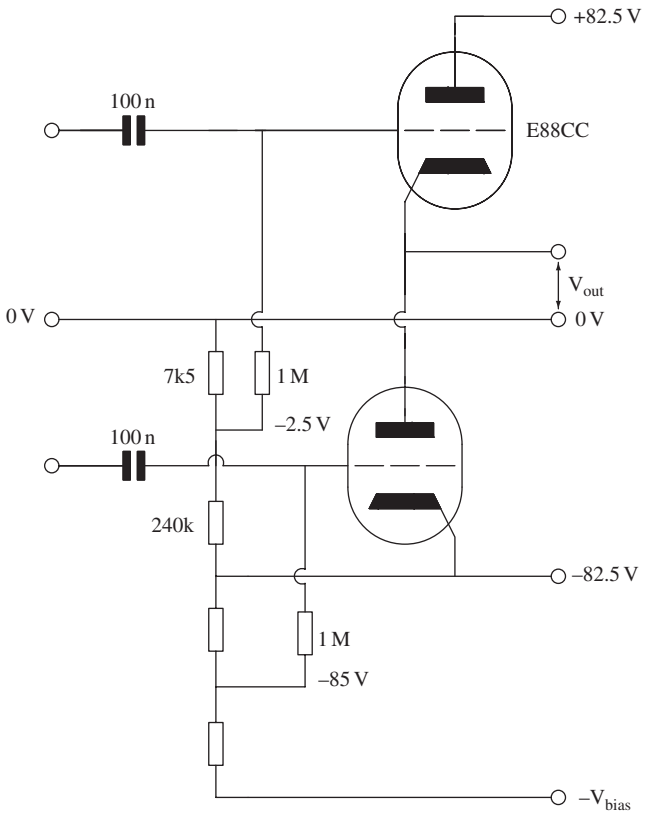


Fig. 2.35 *Push–pull input White cathode follower*

The upper valve no longer has a resistor in its anode circuit, so $r_k = 1/gm$, and this is the anode load of the lower valve. Substituting:

$$A_{v(lower)} = \frac{\mu \cdot \frac{1}{gm}}{\frac{1}{gm} + r_a}$$

Multiplying by *gm*, and simplifying:

$$A_{v(lower)} = \frac{\mu}{\mu + 1}$$

A cathode follower with $R_k = \infty$ would have the same gain, and because the lower valve strives to produce exactly the same signal as a standard

cathode follower if it saw $R_k = \infty$, there is no voltage difference between the two valves, so the upper valve *does* see $R_k = \infty$. However, the input to the lower valve must be inverted, requiring an external phase splitter.

The lower valve no longer reduces the output resistance of the upper valve, since with a gain of 1, it cannot apply feedback to the upper valve, which is why output transformer-less amplifiers need very large amounts of global feedback to bring their output resistance down to a suitable value for damping moving coil loudspeakers.

The μ-follower

This is a design which has attracted considerable interest since its rediscovery a few years ago.[9] (There is nothing new under the sun.) Essentially, it is a common cathode amplifier with an active load. Unlike the cathode follower, where it is arguable whether this is really necessary, the common cathode amplifier can definitely benefit from this sort of treatment. See Fig. 2.36.

The top valve is a self-biased cathode follower which has its input capacitively coupled from the anode of the common cathode lower stage. Since the cathode follower has $A_v \approx 1$, and is non-inverting, the signal at its cathode will be nearly equal to that at the anode of the lower valve. If this is the case, then there will be little, or no, signal voltage across the upper resistors. Little, or no, signal current flows, implying a high resistance active load, or constant current source. The lower valve achieves voltage gain $A_v \approx \mu$, and it produces low distortion (r_a is no longer a factor). As a bonus, we have two output terminals, either the direct output from the lower anode, or the low resistance output from the cathode follower. It should be noted, however, that the high resistance active load actually only operates at AC, since the coupling capacitor forms a high-pass filter in conjunction with the (admittedly high) cathode follower input resistance.

If the upper valve is a constant source (even if only at AC), then we can plot the loadline for the lower valve as a horizontal line. See Fig. 2.37.

This is an example of an *AC loadline*, where the slope of the loadline does not relate to DC conditions, although it must pass through the DC operating point. We can move this line to any operating point that we like. If we choose an anode current of 2 mA, and bias the anode voltage to 80 V this gives $\mu = 32.5$, and so we would expect a gain of ≈ 32.

We now have to determine the operating point for the upper stage. We will supply the compound stage from an HT of 285 V. (When we investigate

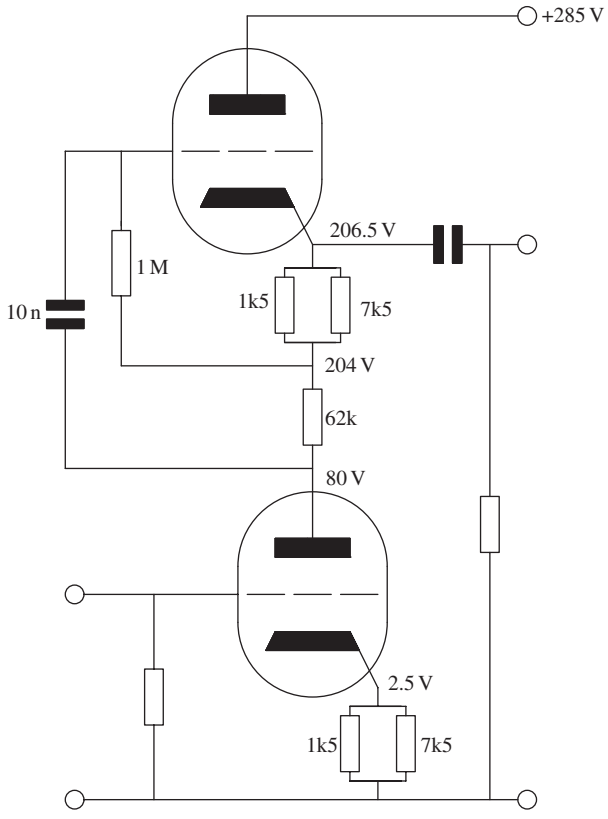


Fig. 2.36 *μ-follower*

power supplies, we will see why 285 V is such a convenient HT voltage.) This leaves 205 V for the upper stage. Since the anode currents are equal, I_a for the upper valve must also be 2 mA. If we now choose an anode voltage for the upper valve (I have chosen 80 V), we can plot the loadline. At $V_a = 0$, we have a current of 3.25 mA, which corresponds to a 63 kΩ total cathode load for the upper valve. V_{gk} for the upper valve is 2.5 V, for $I_a = 2$ mA, we will need a 1.25 kΩ cathode bias resistor. We have now established the DC conditions of the stage.

Once we know the gain of the cathode follower, we can determine the value of the active load that it presents, and find its input resistance, which will enable us to choose an appropriate value for the coupling capacitor.

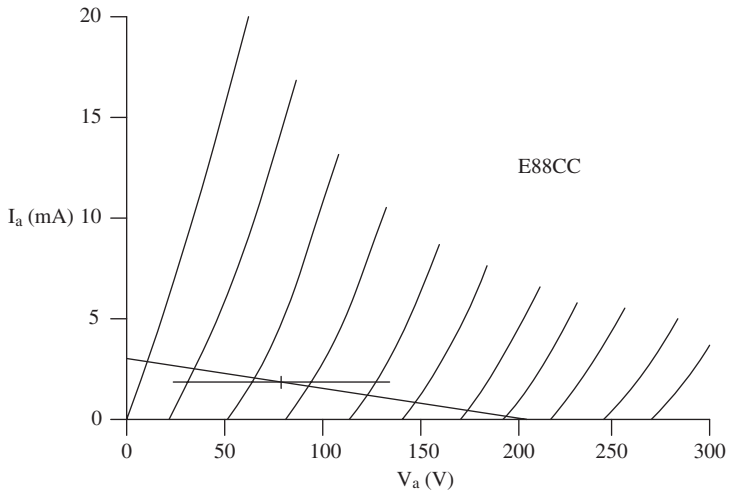


Fig. 2.37 *Operating conditions of lower valve in μ-follower*

From the loadline, the gain before feedback is 29, so the gain of the cathode follower is 29/30, which is 0.97. The lower valve sees an anode load of:

$$r_{\text{load}} = \frac{R_{\text{L}} + R_{\text{k}}}{1 - A}$$

This gives a value of $\approx 2\,\text{M}\Omega$, so our earlier assumptions about the gain and linearity of the lower stage were justified. We can use our earlier formula to determine the input resistance at the grid of the cathode follower:

$$r_{\text{input}} = \frac{R_{\text{g}}}{1 - A \cdot \frac{R_{\text{L}}}{R_{\text{L}} + R_{\text{k}}}}$$

This gives an input resistance of $\approx 19\,\text{M}\Omega$. If we need a 1 Hz cut-off, then 10 nF is perfectly adequate.

The cathode bias resistor for the lower valve was calculated in the normal way. Although the high value of load resistance for the lower valve causes β to be so small that the reduction of gain due to negative feedback is insignificant, we should still include a decoupling capacitor, otherwise the lower anode resistance rises. Increased anode resistance is undesirable because it reduces the anode's ability to drive stray capacitance and shunt sources of noise such as induced hum or resistor noise.

An extremely useful secondary advantage of the μ-follower is its increased immunity to noise on the HT supply, known as *Power Supply Rejection Ratio* (*PSRR*). At the output of any common cathode amplifier, PSRR can be found:

$$\text{PSRR} = \frac{R_L + r_a}{r_a}$$

This is quite simply because r_a forms a potential divider with R_L. For maximum rejection of HT noise and ripple, R_L should be as high as possible compared to r_a. A pentode has $r_a > R_L$ and therefore has *no* rejection of HT noise.

Cathode feedback considerably increases r_a, but does not reduce total gain by a proportionate amount, and therefore destroys HT rejection. In our (bypassed) example, r_a for the lower valve $= 6\,\text{k}\Omega$, and the active load $\approx 2\,\text{M}\Omega$, resulting in 50 dB rejection of HT noise, but removing the bypass would raise r_a for the lower valve to $47\,\text{k}\Omega$, and reduce HT rejection to 33 dB, despite leaving gain relatively unchanged.

Strictly, we should include the loss of the cathode follower in any calculation of gain to the low resistance output ($A_{\text{total}} = \mu \times A_{\text{cathode follower}}$), giving a gain of 31.5 in this instance.

The importance of the AC loadline

Up until now we have tacitly assumed that the input resistance of the following stage had little or no effect on the performance of the preceding stage. This would not be true if we used the anode output of the μ-follower, because the value of the following grid leak (typically $\approx 1\,\text{M}\Omega$) is not merely comparable with the value of R_L, it is actually *less* than R_L, and therefore lowers the effective value of R_L from $\approx 2\,\text{M}\Omega$ to $666\,\text{k}\Omega$. This has a negligible effect on gain, but trebles the distortion, so using the anode output is not recommended.

Whilst $R_g \geq 10R_L$, it is legitimate to ignore its effect on the preceding stage, but once it becomes smaller than this, we should consider drawing an AC loadline to investigate whether it will cause a problem. Stages with active loads *must* take into account the input resistance of the following stage.

An accurate AC loadline is easily drawn. First, we find the AC load, which is usually just the anode load and the following grid leak in parallel. We know that the AC loadline must pass through the operating point, so all we need is a second point. The simplest way to do this is to move a convenient number of squares horizontally (change the voltage by 100 V or

so), and calculate the increase, or decrease, in current through the AC load to give our second point. The line through these points is then the AC loadline, so inspection of this line gives the gain and linearity of the stage *including* the effects of the following load resistor.

Upper valve choice in the μ-follower

There is no reason why the upper valve should be the same as the lower valve in a μ-follower. As a rule of thumb, the AC load resistance seen by the lower valve can be found by:

$$r_L \approx u_{upper} \cdot R_L$$

Maximizing r_L minimizes distortion in the lower valve, but experiment shows that once $r_L \geq 50r_a$, there is no further benefit to be gained and it is more profitable to investigate the distortion produced by the upper valve. Because a cathode follower operates with 100% feedback, increasing μ increases feedback and reduces distortion. However, high μ valves need a higher value of V_a to avoid grid current, reducing available V_a for the lower valve, and lowering maximum voltage swing.

High *gm* is also useful in the upper valve if the stage is to feed a passive equalization network because the resultant low (but changeable) r_{out} is a smaller proportion of the network's series resistance.

The 6C45П single triode has $\mu = 52$ and $gm \approx 20\,mA/V$ at sensible anode currents, but its main advantage as the upper valve is that it can swing V_{gk} close to 0 V without distortion, allowing a high output voltage swing from a given HT voltage.

When triode strapped (g_2, g_3 to a), the D3a pentode is also a good choice as $\mu = 80$, and $gm \approx 20\,mA/V$ is easily achievable even at quite low currents, but significant grid current begins at $V_{gk} \approx -1.1\,V$. The D3a has gold-plated pins and was produced in the era when gold plating meant special quality. Not only does it meet its published specification, but it is very consistent from one sample to another. By contrast, the (Russian) 6C45П generally only just meets the lower limits of its specification and is rather variable, although its anode curves are extremely linear.

Limitations of the μ-follower

Although the μ-follower is an excellent gain stage, it does have limitations. We have seen that it has a low output resistance and low distortion, and it is

therefore tempting to use it as a line stage to drive long cables or low input resistance transistor amplifiers.

However, a low impedance load steepens the AC loadline of the cathode follower upper valve. Although this valve has 100% feedback, the steeper loadline slightly reduces gain and the cathode follower can no longer bootstrap the lower stage's R_L as effectively, so the lower valve sees a reduced load resistance, and distortion in the lower valve rises. Connection of an external load always increases distortion within a μ-follower. As an extreme example, the r_{out} of a 6J5/6J5 μ-follower was tested at 0 dBu. At +28 dBu the stage produced 0.29% THD + N, so it was predicted to produce ≈0.01% THD at 0 dBu. However, when its output resistance was measured by dropping the output from 0 dBu to −6 dBu by loading it with a 720 Ω resistance, THD + N rose to 0.85%.

If a low load resistance must be driven with minimum distortion, the μ-follower can be buffered by a cathode follower. In order to drive the load effectively, the cathode follower should pass ≥10 mA, and the valve should be a frame-grid type with high *gm* and high μ so a 6C45Π or triode-strapped D3a are ideal. The cathode follower is a high impedance load, so it can be direct coupled to the lower output of the μ-follower to avoid the inevitable distortion of the upper valve in the μ-follower. For lowest possible distortion, the cathode follower should have a constant current sink as its load.

A well-designed μ-follower enters overload very suddenly. A 6J5/6J5 μ-follower driven from a 51 kΩ source was driven into grid current, resulting in an output of +38.1 dBu (61.6 V_{RMS}) at 0.87% distortion. A high source resistance causes hard clipping at the onset of grid current, so we should expect a decaying series of odd harmonics. However, because grid current only clips one half cycle, and asymmetry causes even harmonics, we can expect all possible harmonics. See Fig. 2.38.

Dropping the level by 1 dB to +37.1 dBu reduced the distortion to 0.54%, and the higher harmonics have entirely disappeared. See Fig. 2.39.

The Shunt Regulated Push–Pull (SRPP) amplifier[10]

This stage was developed in the early 1950s for use as a power amplifier or modulator in television transmitters, where it was required to drive 1100 V_{pk-pk} into a load of 400 Ω in parallel with 500 pF with low distortion.[11] Far more distortion can be tolerated in video than in audio, and the standards of video at the time were comparatively poor, so 'low distortion' meant ≈2%, and 'negligible distortion' meant <1%.

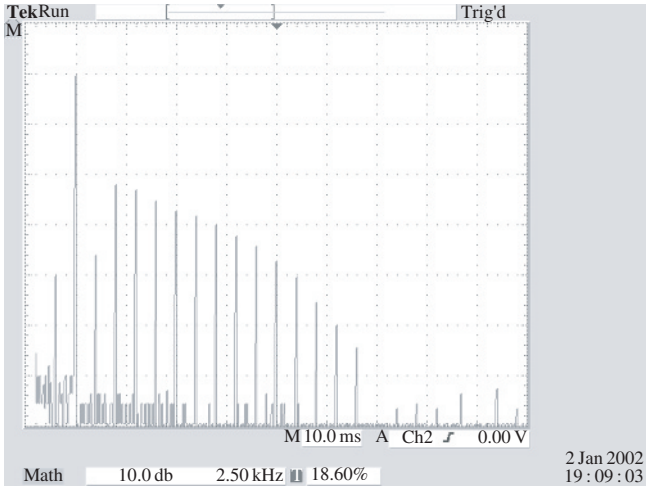


Fig. 2.38 *Distortion spectrum of 6J5/6J5 μ-follower entering grid current*

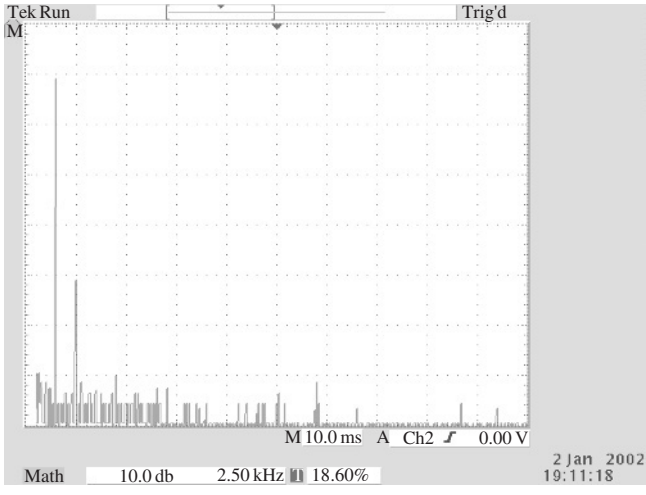


Fig. 2.39 *Distortion spectrum of 6J5/6J5 μ-follower 1 dB below grid current*

Although we are unlikely to use the SRPP in its main application, it is useful to understand the problems that the transmitter engineers faced, and how they were solved. Developing 1100 V_{pk-pk} across 400 Ω wasn't really a problem – it just needed a large valve, but maintaining that voltage across the

500 pF capacitance was. The highest frequency produced by the 405 line 'high definition' system was 3 MHz, and at this frequency, $X_c \approx 100\,\Omega$, requiring considerably more current than the 400 Ω resistance. The obvious solution was to increase the standing current in the stage, but this would have been wasteful of electricity because full amplitude high frequency signals are very rare in real pictures (as opposed to test signals). What was needed was a means of sensing when the extra current was required, and then allowing a second valve to furnish that current. A resistor in series with the output of the lower valve senses the load current, so the voltage across this is used to drive the regulator (upper) valve. Because the regulator valve could typically quadruple the total signal power of the stage without requiring any additional standing current, this stratagem allowed a considerable increase in efficiency – a very important consideration in amplifiers dissipating kW of heat.

The same type of valve is invariably used for both upper and lower sections. See Fig. 2.40.

The same current passes through both valves, so the associated cathode bias resistor R_k is the same. From a DC point of view, the section above the lower anode is identical to that below, so each section sees half the HT voltage. If we draw a vertical line on the anode characteristics at $285\,V/2 = 142.5\,V$,

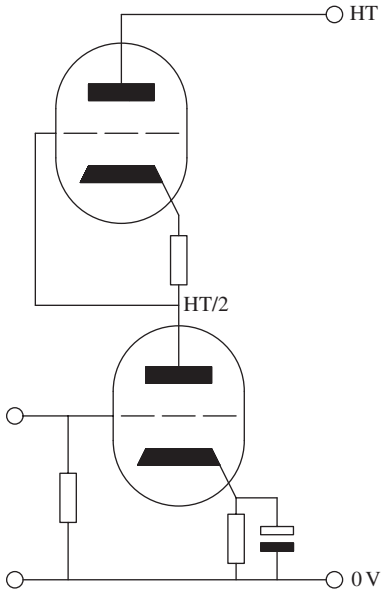


Fig. 2.40 *Shunt Regulated Push–Pull (SRPP) amplifier*

and choose our anode current, this determines the required bias. The −4 V curve crosses 142.5 V at 4.5 mA, so 4 V/4.5 mA = 889 Ω, and a 910 Ω resistor is fine. See Fig. 2.41.

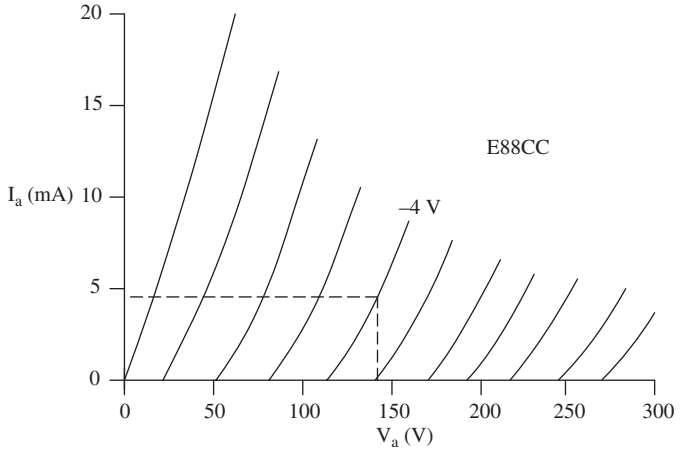


Fig. 2.41 *Choosing SRPP operating current*

Conceivably, differing valves or differing DC conditions could be used for upper (V_2) and lower (V_1) valves, in which case, the full equations derived by Amos and Birkinshaw[12] give the gain of the stage as:

$$A_v = \frac{\mu_1(\mu_2 R_k + r_{a_2})}{r_{a_1} + r_{a_2} + R_k(\mu_2 + 1)}$$

And the output resistance may be found from:

$$r_{out} = \frac{r_{a_2}(R_k + r_{a_1})}{r_{a_1} + r_{a_2} + R_k(\mu_2 + 1)}$$

The SRPP is intermediate between the common cathode amplifier with a resistor as R_L, and the μ-follower with an active R_L, but the low value of upper cathode resistance R_k means that the value of R_L seen by the lower valve is inevitably quite low, implying that the SRPP must have $A_V < \mu$, and significantly increased distortion compared to a μ-follower.

A pair of typical 6J5GTs whose characteristics had previously been measured were set up as an SRPP:

$$\mu = 21$$
$$gm = 2.95\,\text{mA/V}$$
$$r_a = 7.11\,\text{k}\Omega$$

The equations predicted $A_v = 14.3$ and $r_{out} = 2.3\,\text{k}\Omega$. Measurement found $A_v = 13.5$ and $r_{out} = 2.3\,\text{k}\Omega$. The SRPP was compared with the μ-follower by swapping the same valves between stages, and testing with identical DC conditions for all valves. Unsurprisingly, given its heritage, not only did the SRPP deliver a significantly higher output voltage swing than the μ-follower, but the μ-follower required a higher HT voltage because it wastes HT across its additional 10 kΩ R_k. See Fig. 2.42.

At an output of +28 dBu (19.5 V_{RMS}), the μ-follower produced 0.24% THD + N, but the SRPP produced 1.32%, an increase of 15 dB. As predicted, the SRPP produces significant distortion, and although this falls with level, it is still rather high for use in a pre-amplifier gain stage. The effect of HT voltage on distortion at +28 dBu was also investigated. See Fig. 2.43.

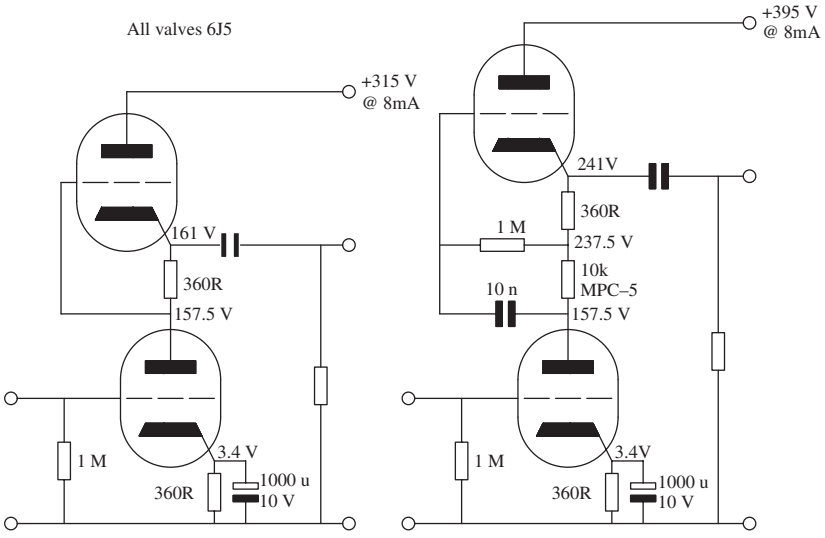


Fig. 2.42 *Circuits used for comparison of SRPP and μ-follower*

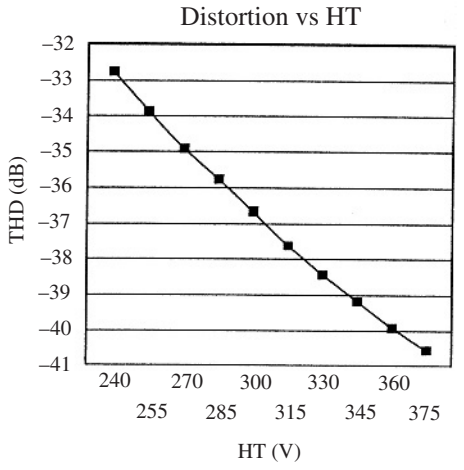


Fig. 2.43 *THD vs HT voltage for 6J5/6J5 SRPP at +28 dBu*

Although the SRPP seems a poor choice compared to the μ-follower, it does have the advantage that it is DC coupled internally (the μ-follower needs a coupling capacitor to the upper valve), and it is therefore immune to blocking (see Chapter 3).

The β-follower

The β-follower[13] seeks to exploit the advantages of the μ-follower with the efficiency and DC coupling of the SRPP stage. See Fig. 2.44.

Replacing the cathode bias resistor with a bipolar transistor allows the large (perhaps 10 kΩ) R_L to be discarded, reducing wastage of HT, and allowing the two valves to be DC coupled.

Bipolar transistors are usually treated as if their output characteristics are constant current, which implies horizontal output curves, but real transistors have curves that slope slightly. See Fig. 2.45.

Looking into the anode, a valve multiplies R_k by μ, similarly, a bipolar transistor multiplies any resistance in the emitter circuit by β, or h_{fe}, so the curves can be flattened by adding an emitter resistor. Since h_{fe} for a small-signal transistor is likely to be ≈400, a 100 Ω resistor in the emitter makes the output resistance $1/h_{oe} \approx 40$ kΩ. The cathode follower then multiplies

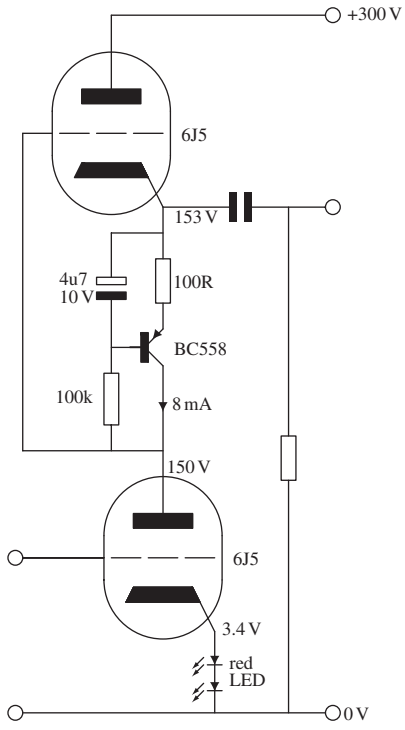


Fig. 2.44 *β-follower*

this resistance by its μ, perhaps 20, to give $R_L \approx 8\,M\Omega$, which is even better than a μ-follower can achieve.

The β-follower easily achieves $r_L \geq 50r_a$, even with a low μ upper valve, so the upper valve must be chosen for minimum distortion, otherwise it will compromise the excellent performance of the lower valve.

The β-follower is an excellent test bed for determining irreducible distortion. If the lower valve is fed from $r_s \approx 0$, and loaded by $r_L \approx \infty$, then the remaining distortion is due to errors in valve geometry, such as uneven grid winding. The 6J5/6J5 β-follower in Fig. 2.44 gave distortion performance that challenged the author's test equipment, with only the 2nd harmonic being reliably measurable at −55 dB below the fundamental at an output level of +28 dBu – all other harmonics were better than −100 dB!

In theory, we could replace the bipolar transistor and its associated components by a depletion-mode JFET. If the gate is connected directly to the source, a typical 2SK147 becomes a ≈9 mA constant current source. However, r_d (the output resistance looking into the drain) is typically

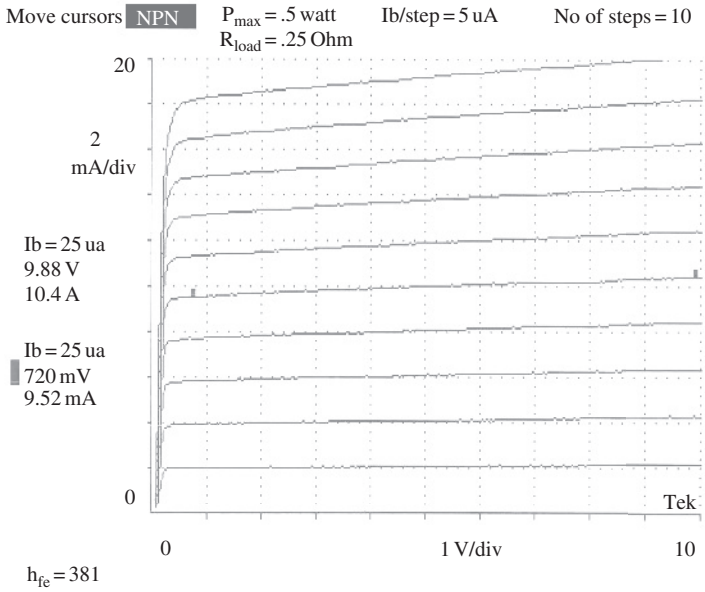


Fig. 2.45 *I_c vs V_{ce} for BC549 NPN transistor*

<10 kΩ, so it is not as effective at reducing distortion as the β-follower. See Fig. 2.46.

The differential pair

All of the circuits that we have so far studied have been *single ended*, which is to say that they have only one output. (The μ-follower type circuits were single ended, because although they had two outputs, they were of the same polarity.)

By contrast, the *differential pair* has two inputs, and amplifies the *difference* between them, to provide two outputs, one inverted with respect to the other – this makes the differential, or *long-tailed* pair[14] a very useful stage.

A differential pair can be made using the basic common cathode triode amplifier, or with cascodes. (The μ-follower is not suitable because differential pairs attempt to exploit the normally large ratio between R_L and R_k.) For simplicity, we will analyse the differential pair using the basic common cathode triode amplifier. See Fig. 2.47.

The circuit consists of two identical triodes, often in the same envelope, with their cathodes tied together, passing anode current to ground via a constant current sink, and each driving equal value anode load resistors.

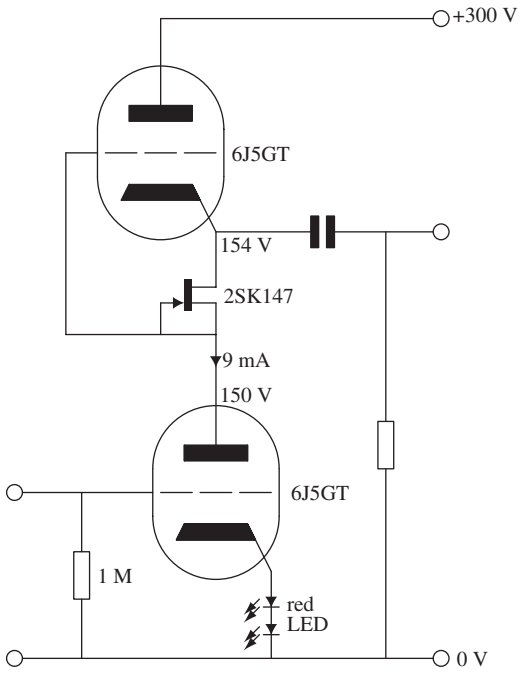


Fig. 2.46 *μ-follower with 2SK147 JFET constant current source*

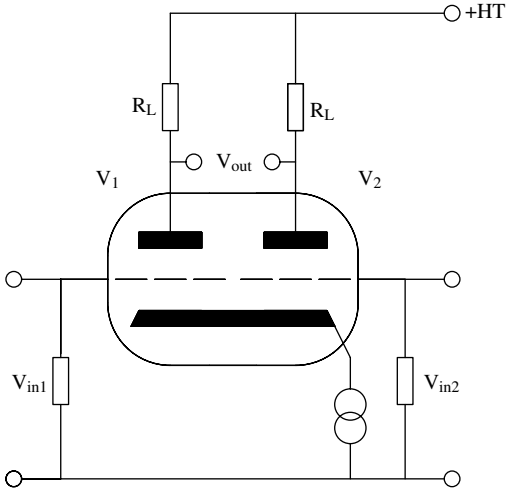


Fig. 2.47 *The differential or long-tailed pair*

Suppose that we apply an input signal such that the voltage on the anode of V_1 rises by 1 V. The current through V_1 must therefore have fallen, but since both valves are sitting on a constant current sink, this can only occur if the current through V_2 has *risen* by an *equal* amount. Since the anode load resistors are equal, it follows that the voltage on the anode of V_2 must have fallen by 1 V.

The outputs of the two anodes are equal in voltage, but one is *inverted* with respect to the other.

Returning to the inputs: if we short circuit g_{V_2} to ground, and apply a sine wave to g_{V_1}, then the cathode will 'follow' that signal, because, ignoring the anode loads, the circuit is a cathode follower. This means that V_2 is driven by its *cathode*, an amplified sine wave appears on its anode, and therefore an equal and opposite signal appears on the anode of V_1. The argument works in the same manner for a signal applied to g_{V_2}.

Gain of the differential pair

When driven by a signal connected between the two grids, the gain of the differential pair is identical to that of a standard common cathode stage, but the output voltage is found *between* the anodes of the stage. Therefore, if we look between one anode and ground, we only see half of the output voltage, and the gain appears to be halved.

If we use the differential pair as a phase splitter, and apply the same input voltage as before between one grid and ground, instead of each grid seeing half the input voltage, one sees the entire input voltage and the other none. Because the voltage difference between the two grids is the same, the gain remains the same.

Output resistance of the differential pair

Provided that the output of the differential pair is not unbalanced *in any way*, r_{out} at each terminal is identical to that of a simple common cathode amplifier ($r_a \| R_L$).

However, if only one output is loaded, the output resistance rises considerably. Working backwards from the path to ground (HT supply) via the first R_L, we see:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

But we now also see a path R_k to ground (0 V), which is in parallel with r_k:

$$r'_k = \frac{R_k \cdot \frac{(R_L + r_a)}{\mu + 1}}{R_k + \frac{(R_L + r_a)}{\mu + 1}}$$

Multiplying through by $(\mu + 1)$:

$$r'_k = \frac{R_k(R_L + r_a)}{R_k(\mu + 1) + R_L + r_a}$$

Looking down through the second anode, we see r_a in series with this multiplied by a factor of $(\mu + 1)$:

$$r'_a = r_a + \frac{R_k(\mu + 1)(R_L + r_a)}{R_k(\mu + 1) + R_L + r_a}$$

If we divide by $R_k(\mu + 1)$, we obtain:

$$r'_a = r_a + \frac{R_L + r_a}{1 + \frac{R_L + r_a}{R_k(\mu + 1)}}$$

As R_k tends to ∞, the right-hand term on the bottom line reduces to zero, resulting in a maximum value of r'_a:

$$r'_a \approx R_L + 2r_a$$

This high value of r_a will become significant when we investigate the PSRR of the differential pair.

If $R_L >> r_a$, then the output resistance (only one terminal loaded) is:

$$r_{out} \approx \frac{R_L(R_L + 2r_a)}{2(R_L + r_a)} \approx \frac{R_L}{2}$$

Common-Mode Rejection Ratio (CMRR)

If we now apply +1 V to *both* grids, the cathode voltage simply rises by 1 V, the cathode current remains constant, and the anode voltages do not change,

because we have not modulated V_{gk}. The amplifier only responds to differences between the inputs, or *differential* signals. Applying the same signal to both grids is known as a *common-mode* signal.

This property of rejecting common-mode signals is significant, since it implies that the circuit can reject hum on power supplies, or common-mode hum on the input signal, so we will investigate it further.

The signal at each output can be expressed in terms of currents using Ohm's law:

$$v_{out(1)} = i_1 \cdot R_{L(1)}$$
$$v_{out(2)} = i_2 \cdot R_{L(2)}$$

Each output will be an exact inverted replica of the other *if* $i_1 = i_2$, provided that the two load resistors are equal. There are two main ways in which this nirvana may be eroded.

- If signal current is lost through an additional path to ground. A signal current i_1 flowing down V_1 out of its cathode must split, with some current being lost down R_k, and the remainder flowing up into the cathode of V_2, to become i_2. However, if $R_k = \infty$, no current can flow down R_k, and $i_1 = i_2$. If $\mu_1 = \mu_2$, and $R_{L(1)} = R_{L(2)}$ then:

$$\text{CMRR} \approx \frac{\mu R_k}{R_L + r_a}$$

  This indicates that we should use high μ valves, and maximize the *ratio* of R_k to R_L. As an example, the second stage of the balanced RIAA pre-amplifier (Chapter 7) has an EF184 constant current source ($r_{sink} = R_k \approx 1\,\text{M}\Omega$), E88CC differential pair ($\mu = 32$), and $R_L = 47\,\text{k}\Omega$, so CMRR $\approx$ 57 dB.
- Results from the above equation will be degraded if $\mu_1 \neq \mu_2$, or if $R_{L(1)} \neq R_{L(2)}$, or a combination of the two. With the easy availability of low cost, accurate digital multimeters, mismatching of load resistors is avoidable, but matching the valves is harder. If $\mu_1 \neq \mu_2$ then:

$$\text{CMRR} \propto \frac{\mu_1 \cdot \mu_2}{\mu_1 - \mu_2}$$

Which indicates that high μ valves are still desirable, but that matching is important.

Because the simple equation for CMRR ignores mismatched valves, mismatched load resistors, and stray capacitances, any predictions of CMRR > 60 dB should be treated with caution. Nevertheless, it is handy for checking that your tail resistance R_k is sufficiently high to ensure that predicted CMRR > 40 dB, since 40 dB is an easily achievable CMRR in practice.

Power Supply Rejection Ratio (PSRR)

Since hum and noise on the power supply line is a common-mode signal, it must also be attenuated by the CMRR. We might also expect the potential divider formed by r_a and R_L to give significant additional attenuation. However, at either terminal, the only path to ground is via the other anode and R_L up to the HT supply, which is exactly the scenario we investigated when determining r_{out} with only one output loaded, therefore:

$$r_a' \approx R_L + 2\,r_a$$

And the attenuation of power supply noise (solely due to potential divider action) is thus:

$$\text{Attenuation} = \frac{R_L + 2r_a}{2(R_L + r_a)} \leq 6\,\text{dB}$$

If $R_L >> r_a$, we achieve the maximum attenuation of 6 dB! Our previous example ($R_L = 47\,k\Omega$, $r_a = 4.95\,k\Omega$) attenuates by 5.2 dB, and together with the 57 dB due to CMRR, PSRR = 62 dB.

It is now well worth comparing the PSRR of the common cathode stage, μ-follower, and differential pair (same DC conditions for the amplifying valve).

Stage	**PSRR**
Common cathode ($R_L = 47\,k\Omega$)	20 dB
μ-follower ($r_L \approx 740\,k\Omega$)	44 dB
Differential pair ($r_{sink} \approx 1\,M\Omega$)	62 dB

The differential pair is the best, and will remain the best, since an improved constant current source for the μ-follower could be adapted to become an improved constant current sink for the differential pair.

Knowing the PSRR enables us to design power supplies correctly because it gives an indication of the allowable hum on the HT supply.

As an example, the second stage of the balanced pre-amplifier (see Chapter 7) needs the 100 Hz power supply hum to be 100 dB quieter than the maximum expected audio signal. At this point, the signal has not received RIAA 3180 μs/318 μs correction, so the level at 100 Hz is 13 dB lower than at 1 kHz. However, peak levels from LP are +12 dB compared to the 5 cms^{-1} line-up level, so the maximum audio signal at 100 Hz is 1 dB lower than the 1 kHz calculated signal level at the anode (2.2 V_{RMS}) = 2 V. We want 100 dB signal/hum, but because 62 dB of this will be provided by PSRR, we only need the hum on the power supply to be 38 dB quieter than 2 V, so we could tolerate 25 mV of hum on the power supply – which is easily achievable.

Semiconductor constant current sinks

The differential pair demonstrated the need for constant current sinks, but the pentode constant current sink is profligate with HT voltage (although it is a very good sink), and a differential pair with grids at ground potential would require a subsidiary negative supply for the sink of −100 V. This is often undesirable, so a solution is needed.

Unlike the original valve designers, we are in the fortunate position of being able to use transistors, and even op-amps, if we consider them to be necessary. This is a perfect example of where a transistor or two can be very helpful.

The simplest form of a transistor constant current sink is very similar to our triode version. The red LED sets a constant potential of ≈1.7 V on the base of the transistor. V_{be} is ≈0.7 V, so the emitter resistor has 1 V held across it. If we need to sink 5 mA, we would use a 200 Ω sense resistor. The AC resistance looking into the collector is:

$$r_{out} = R_E \cdot h_{fe} + 1/h_{oe}$$

In this instance, a BC549 ($h_{fe} \approx 400$, $1/h_{oe} \approx 12\,k\Omega$) gives $r_{out} \approx$ 92 kΩ. Note that an expensive 2 W resistor is required to bias the LED. See Fig. 2.48a.

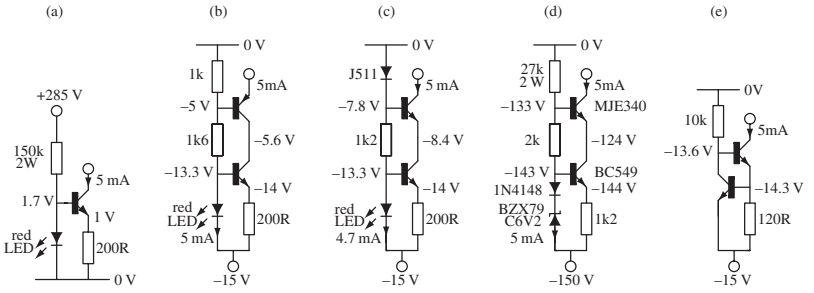


Fig. 2.48 *Semiconductor constant current sources*

The simple circuit can easily be improved upon, and since silicon is cheap, it seems worthwhile to do so. There are two problems to be addressed. First, the transistor needs $V_{CE} > 0.5\,V$ for it to operate as a constant current sink, which is uncomfortably close to typical bias voltages for high μ valves such as the ECC83. Second, 92 kΩ output resistance is not especially high, and we can do much better.

A transistor cascode is broadly similar to a pentode, but a practical circuit requires a negative supply. However, this may not be a problem in a power amplifier, because there is often a negative bias supply for the output valves that we can use. (Even though the bias winding normally supplies <1 mA, wire rated for 1 mA is very fragile, so transformer manufacturers typically use thicker wire, allowing us to draw 10 mA from this winding, and the increase in total transformer VA loading is usually negligible.)

The cascode constant current sink has much higher output resistance than a single transistor constant current sink:

$$r_{out} = R_E \cdot h_{fe(upper)} \cdot h_{fe(lower)} + 1/h_{oe(upper)}$$

The AC output resistance of the initial design has been multiplied by the h_{fe} of the second transistor which improves it from ≈92 kΩ to ≈32 MΩ, so the value of $1/h_{oe}$ is now negligible. However, a more practical advantage is that the negative supply allows the output port to be taken down to 0 V without linearity problems. High frequency stability is excellent. See Fig. 2.48b.

As shown, the cascode current source is relatively sensitive to hum and noise on the negative supply because of current changes through the voltage reference. This sensitivity can be greatly reduced by modifying the circuit to include a current regulator diode in the chain that feeds the voltage reference. See Fig. 2.48c.

The cascode constant current sink can be adapted to withstand a larger voltage simply by substituting the transistor that feeds the load for a higher voltage type. This slightly lowers r_{out}, because the higher voltage transistor inevitably has a lower h_{fe}, but because we now have volts to spare, most of this loss can be recovered by setting a higher reference voltage, allowing a higher value of R_E. Unfortunately, if a power transistor is required, its higher output capacitance degrades performance at high frequencies. The 1N4148 diode compensates for variation of the lower transistor's V_{be} due to temperature, but requires all component values to be recalculated. See Fig. 2.48d.

The 'ring of two' circuit works by holding 0.7 V across the 120 Ω sense resistor. If that voltage rises, due to increased current through the resistor, T_1 turns on harder, which causes the base voltage of T_2 to fall. T_2 begins to turn off and so the current through the 120 Ω resistor, and therefore the sink current, is held constant. Because this circuit uses feedback applied over two transistors, there is a possibility of oscillation at high frequencies due to stray capacitances. See Fig. 2.48e.

Using transistors as active loads for valves

All the previous sink circuits can be mirrored about 0 V and PNP transistors substituted for NPN. If the circuit is then connected to the HT supply, they become constant current sources allowing a triode to achieve $A_v = \mu$. More significantly, they permit a valve to achieve low distortion from a low HT voltage.

As an example, in common with all high μ valves, the ECC83 needs considerable V_a before it can be biased out of grid current, 150 V is typical. As a general rule of thumb, $R_L > 2r_a$, and as $r_a \approx 75\,k\Omega$ for the ECC83, we might use $R_L = 150\,k\Omega$. If $I_a = 0.7\,mA$, we would drop 105 V across R_L, so we would need 255 V of HT. But we might only require the stage to produce an output swing of $5\,V_{pk-pk}$, so most of the HT is wasted. If we replace the 150 kΩ resistor with a constant current source the valve sees a much higher value of R_L, and we can set the HT voltage independently to accommodate the maximum required output swing. See Fig. 2.49.

In Fig. 2.49, the concept of operating a high μ valve from a low HT was taken to the extreme because the author needed a high gain differential pair stage (ECC83: $\mu = 100$), but only had 150 V of positive HT available. Note that high voltage transistors are required to withstand either anode swinging towards 0 V.

Although Zener diodes are normally bypassed to reduce noise, the noise generated by both Zener diodes is common mode, and is therefore rejected by

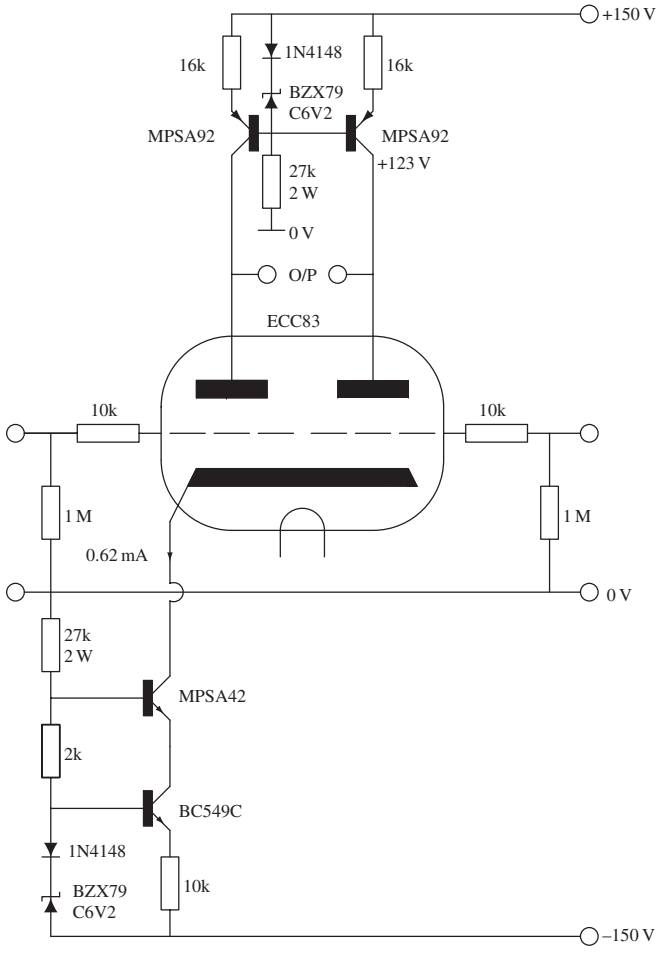


Fig. 2.49 *Semiconductor anode loads. Note: Fine adjustment of tail current sets anode voltages*

the next (differential) stage. On test, the circuit achieved the required differential swing of $7\,V_{pk-pk}$ at 1 kHz with only 0.04% distortion.

A cascode greatly increases r_{out}, flattening the loadline, and reducing distortion in the valve. If we wanted to maximize output swing and minimize distortion, we might operate a 7N7 (Loctal equivalent to 6SN7) at $I_a = 8\,mA$ because μ becomes more nearly constant when $I_a > 6\,mA$. We assume that our cascode will provide a horizontal loadline, so we plot this at 8 mA. See Fig. 2.50.

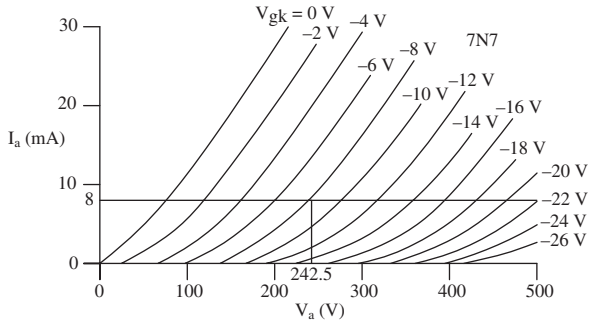


Fig. 2.50 *7N7 with constant current 8 mA loadline*

Normally, as V_a rises, we have to consider the cramping of curves as I_a falls and cut-off is approached, but I_a is now a constant, and the only limit to positive swing is that the cascode requires sufficient voltage to operate correctly. 15 V is quite adequate for the cascode, so a 400 V HT would allow V_a to swing to 385 V. Looking in the opposite direction along the 8 mA loadline, grid current is likely to begin at ≈100 V. The maximum possible swing is therefore $385\,V - 100\,V = 285\,V_{pk-pk} \approx 100\,V_{RMS}$.

Although the cascode forces $I_a = 8\,mA$, we must adjust valve bias to set V_a. At just over maximum swing we should clip positive and negative half cycles equally, so the operating point should be halfway between the maximum and minimum permissible anode voltages – which is their average:

$$V_a = \frac{V_{max.} + V_{min.}}{2} = \frac{385 + 100}{2} = 242.5\,V$$

Looking at the curves, we see that $V_{gk} \approx 8\,V$ is required to set V_a correctly, and this can be provided by an 8.2 V Zener diode. See Fig. 2.51.

Since the stage is intended to swing large voltages, noise is not a problem, so it is not necessary to bypass the Zener diode with a capacitor.

If there is 242.5 V across the valve, then there is 147.5 V across the lower transistor, so it must dissipate 1.18 W when quiescent. When V_a swings to 100 V, the transistor must withstand 285 V at 8 mA, so it momentarily dissipates 2.28 W, and it might seem that this is the required rating of the transistor. However, the flat loadline has reduced distortion almost to zero, so the positive and negative swings are equal, and the average power dissipated in the transistor over one cycle of audio is equal to the quiescent power.

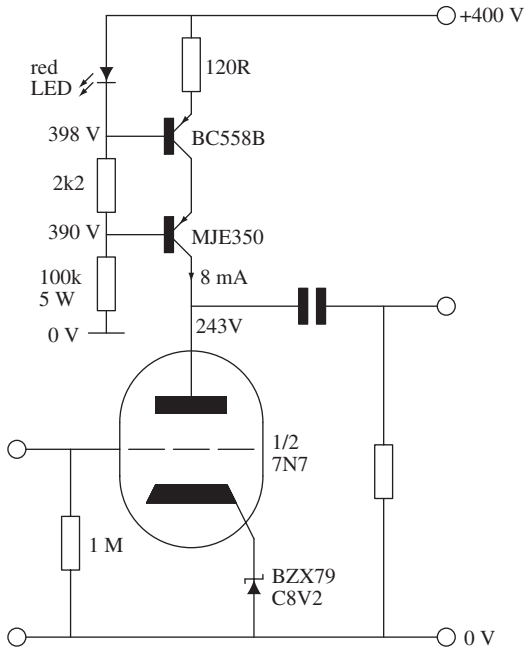


Fig. 2.51 *Cascoded semiconductor anode load*

As before, it is only the transistor nearest the valve that must be able to withstand high voltages and dissipate significant power, so the additional transistor can be as fast and fragile as we like.

Optimizing r_{out} by choice of transistor type

The table compares transistors that are useful in support circuitry for valves.

Output resistance at low frequencies is partly determined by $1/h_{oe}$, but is dominated by h_{FE}, since any resistance in the emitter circuit is multiplied by h_{FE}. Impedance at high frequencies is shunted by the capacitance seen at the collector of the transistor, which will partly be determined by strays, but also by the transistor itself. In general, high voltage/high power transistors have a larger silicon die area, and greater capacitance, which is reflected in their lower f_T. Additionally, f_T varies significantly with I_C, and operating a transistor below its optimum I_C could reduce f_T by a factor of five. If in doubt, download the transistor's data sheet from the Internet – all the semiconductor manufacturers have excellent websites. As a consequence of

these considerations, a small cascode constant current sink would ideally use two BC549, or if low output capacitance (≈ 0.5 pF excluding strays) was essential and sufficient voltage was available, three BFR90.

		$V_{CE(MAX.)}$	$I_{C(MAX.)}$	$P_{MAX.}$	f_T	$h_{FE(MIN.)}$	$1/h_{oe(TYP.)}$
BFR90	NPN	15 V	25 mA	300 mW	5 GHz	40	5 kΩ
BC549	NPN	30 V	100 mA	500 mW	300 MHz	200	12 kΩ
BC558B	PNP				200 MHz	220	6 kΩ
2N3904	NPN	40 V	200 mA	500 mW	250 MHz	100	15 kΩ
2N3906	PNP			625 mW			5 kΩ
MPSA42	NPN	300 V	500 mA	625 mW	50 MHz	40	50 kΩ
MPSA92	PNP						35 kΩ
MJE340	NPN	300 V	500 mA	20 W	10 MHz	30	150 kΩ
MJE350	PNP				4 MHz		50 kΩ

$V_{CE(MAX.)}$: the maximum allowable voltage between collector and emitter. (There are various ways of specifying this limit, so unless you know your precise circuit conditions, it is wise not to exceed 2/3 V_{CE}.)

$I_{C(MAX.)}$: the maximum allowable collector current.

$P_{MAX.}$: the maximum allowable power dissipation in the device ($P = I_C \times V_{CE}$).

f_T: AC current gain h_{fe} falls with frequency. At f_T, $h_{fe} = 1$; this is known as the transition frequency.

$h_{FE(MIN.)}$: minimum DC current gain from base to collector. (The author's measurements suggest that h_{FE} is generally double the manufacturer's specified minimum value at the typical currents required by valves.)

$1/h_{oe(TYP.)}$: this is the typical AC resistance (equivalent to R_a) seen looking into the collector. It is very rarely specified by manufacturers, so these figures were measured on a curve tracer at $V_{CE} = 10$ V, $I_C = 10$ mA to allow comparison between the types. PNP transistors tend to have a lower Early voltage than their NPN counterparts, so $1/h_{oe}$ is lower *and* falls faster at higher currents.

Any bipolar transistor needs a minimum V_{CE} for it to operate linearly. For a low voltage transistor at currents ≤ 30 mA, ≈ 1 V is sufficient, but higher currents may require 2 V. See Fig. 2.52.

High voltage transistors such as MPSA42 or MJE340 may also require $V_{CE} > 2$ V. A constant current sink in a differential pair operating as a phase splitter has half the input signal across it, so this point can become significant. In a cascode sink, the lower transistor has hardly any AC signal across it, so it can operate with only 2–3 V_{DC}, leaving the remaining DC for the upper transistor, which supports the bulk of the AC signal.

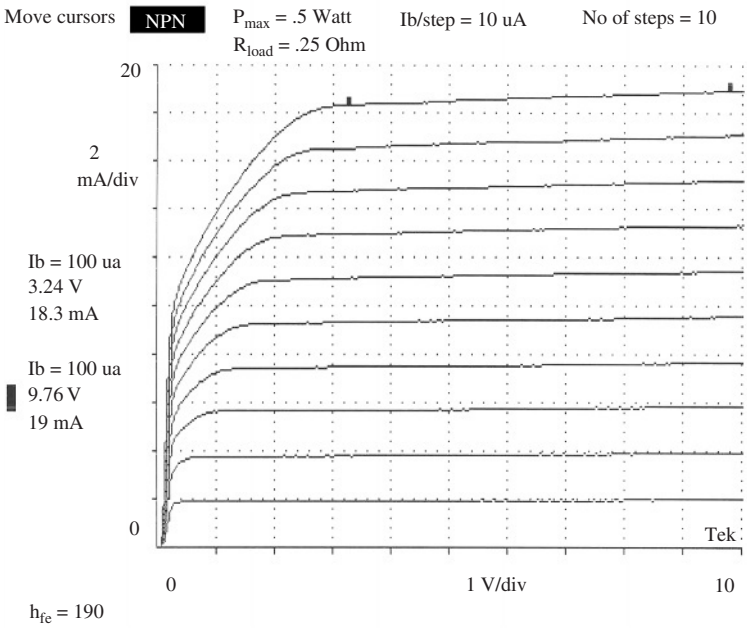


Fig. 2.52 *I_C vs V_{CE} for 2N3904 transistor showing minimum V_{CE} required*

The LM334Z IC constant current sink

Another possibility for a constant current sink is to use a purpose designed programmable constant current sink such as the LM334Z. This device does not need any subsidiary supplies, and operates correctly with only 1.2 V across it, but $I_{(max.)} = 10$ mA. The typical application for this device is in the tail of a differential pair, but stability can be a problem if the differential pair is momentarily driven into overload, since the supply voltage for the internal operational amplifier may then fall so low that it no longer operates linearly, and it oscillates. You have been warned.

References

1 Mullard (1969) Book 2 part 1: Receiving valves and television picture tubes, General Operational Recommendations, p. 17.
2 Gewartowski, J. W. and Watson, H. A. (1965) *Principles of Electron Tubes*. Van Nostrand, pp. 112, 121.

3 British Patent No. 145,421 (1921) Improved vacuum electric discharge apparatus having an incandescent cathode. Siemens & Halske Aktiengesellschaft.
4 British Patent No. 287,958 (1928) Improvements in or relating to circuit arrangements and discharge tubes for amplifying electric oscillations. NV Philips' Gloeilampenfabrieken.
5 British Patent No. 448,421 (1934) Improvements in and relating to thermionic valve circuits. Alan Dower Blumlein.
6 Puckle, O. S. (1943) *Time Bases*, 1st ed. Chapman & Hall p. 92.
7 British Patent No. 564,250 (1940) Improvements in or relating to thermionic valve amplifier circuit arrangements. Eric White.
8 Amos, S. W. and Birkinshaw, D. C. (1958) *Television Engineering: Principles and Practice. Volume 4: General Circuit Techniques*. Iliffe, p. 264.
9 Kimmel, Alan (1993) The mu stage, *Glass Audio*, Vol. 5, No. 2, p. 12.
10 US Patent No. 2,310,342 (1943) Balanced direct and alternating current amplifiers, Radio Corporation of America.
11 Cooper, V. J. (1951) Shunt-regulated amplifiers, *Wireless Engineer*, May, pp. 132–145.
12 Amos, S. W. and Birkinshaw, D. C. (1958) *Television Engineering: Principles and Practice. Volume 4: General Circuit Techniques*. Iliffe, p. 250.
13 Vilá, Ignacio (1998) The beta follower amplifier, *Glass Audio*, Vol. 10, No. 4, p. 1.
14 British Patent No. 482,740 (1936) Improvements in or relating to thermionic valve amplifying circuit arrangements. Alan Dower Blumlein.

Further reading

Agger, L. T. (1955) *Principles of Electronics*, 2nd ed. Macmillan.
Alexander Robert Charles (1999) *The Inventor of Stereo: The Life and Works of Alan Dower Blumlein*. Focal.
Amos, S. W. and Birkinshaw, D. C. (1956) *Television Engineering: Principles and Practice. Volume 2: Video-frequency Amplification*. Iliffe & Sons.
British Standard Code of Practice CP1005 (1962) *The use of electronic valves*.

Burns, Russell (2000) *The Life and Times of A. D. Blumlein*. The Institution of Electrical Engineers.
Chaffee, E. Leon (1933) *Theory of Thermionic Vacuum Tubes*. McGraw-Hill.
Eastman, A. V. (1949) *Fundamentals of Vacuum Tubes*, 3rd ed. McGraw-Hill.
Langford-Smith (1953) *Radio Designers Handbook*, 4th ed. Iliffe. reprinted (1999).
Radio Society of Great Britain (1976) *Radio Communications Handbook*, 5th ed.
Reich, Herbert J. (1939) *Theory and Applications of Electron Tubes*. McGraw-Hill.
Reich, Herbert J. (1941) *Principles of Electron Tubes*. Now reprinted by Audio Amateur Press (1995).
Ryder, J. D. (1964) *Electronic Fundamentals and Applications*, 3rd ed. Prentice-Hall.
Terman, F. E. (1955) *Electronic and Radio Engineering*, 4th ed. McGraw-Hill.
Valley and Wallman (1948) *Vacuum Tube Amplifiers*. McGraw-Hill. Now reprinted by Audio Amateur Press (2000).

3

Distortion

In this chapter, we will investigate distortion, but begin by looking at the fundamentals of distortion measurement. Even designers of test equipment would probably concede that this is not the sexiest of topics, but unless we understand how errors can creep into our measurements, we will not have the confidence to compare the results of one measurement with another, leaving us unable to test and improve our designs.

Defining distortion

Although we may glibly use the word 'distortion' when talking about amplifiers, there are actually two distinct types of distortion.

Linear distortions do not change with amplitude. If we consider the transfer characteristic of a device producing linear distortion, it is a straight line – hence the term linear distortion. See Fig. 3.1a.

Although a device causing linear distortion changes the shape of the waveform, there are no *additional* frequencies at the output of the device. Linear distortion typically causes errors in the amplitude versus frequency response – and this is the way that it is usually assessed. However, it is perfectly possible to change the shape of a waveform without changing the amplitude vs frequency response by distorting the *time* at which frequencies arrive. Loudspeaker crossover systems and analogue tape machines are particularly good at this feat. The shape of a square wave's leading edge is particularly sensitive to timing errors, so an oscilloscope quickly reveals problems. Alternatively, timing errors between sine waves of differing frequencies can be assessed by plotting phase versus frequency response on linear graph paper. Deviations from a straight line imply delay errors – hence the term *linear phase* for an ideal device.

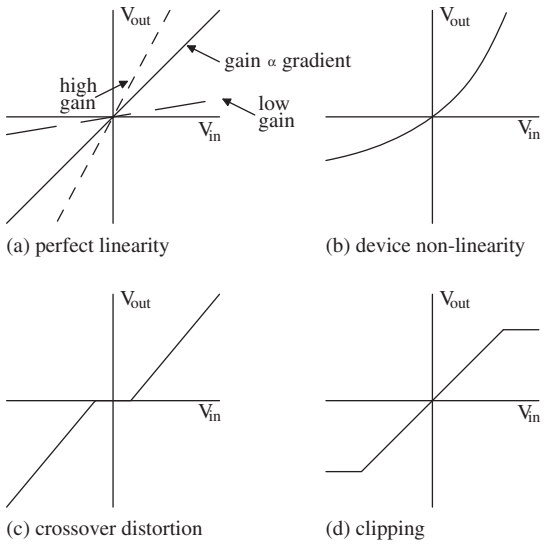


Fig. 3.1 *Transfer characteristics and the distortion they produce*

Unsurprisingly, the transfer characteristic for *non-linear* distortion is not a straight line, and a device causing non-linear distortion has frequencies at the output that were not present at the input. See Fig. 3.1b, c, d.

Measuring non-linear distortion

We can assess the linearity of a device in two fundamental ways:

- We plot the transfer characteristic directly. Since the definition of non-linear distortion was that the transfer characteristic should deviate from a straight line, we could measure the deviations. In practice, this is a poor method because the small deviations required by high quality audio make it difficult to keep measurement errors below the deviations.
- We look for frequencies at the output of the device that were not present at the input. This is a very sensitive and easily applied test, so there are two common variations on this theme.

The simplest expression of the second test is to apply a single sine wave to the device. At the output of the device, we expect to see a single sine wave.

However, if the device produces non-linear distortion, there will also be harmonics of the original sine wave. The test is popular because removing the original sine wave at the output is easy, leaving only the harmonics – which can then be measured either individually or collectively as Total Harmonic Distortion (THD).

A more complex method is to apply *two* sine waves to the device. Again, we should only see these two frequencies at the output, but a device producing non-linear distortion causes the two frequencies to modulate each other, producing sum and difference frequencies. Intermodulation distortion measurement is popular with RF engineers because it is easy to tune to each sum and difference frequency and measure its amplitude, but this is not quite so easily done at audio frequencies.

It is most important to realize that measuring harmonic distortion is no more 'correct' than measuring intermodulation distortion, or vice versa. Both forms of measurement simply reflect the same non-linearity in the device's transfer characteristic. What *is* important is how the measurement is made and how the results are interpreted.

Distortion measurement and interpretation

In an ideal world, everybody would make their distortion measurement in the same way, with the same equipment, and interpret their results identically. All results would then be comparable, allowing us to state that device 'A' was better than device 'B'.

In practice, there are many different measurement techniques. For example, the intermodulation distortion measurement requires two (or more) frequencies. Which frequencies should be chosen, and what should their relative amplitudes be? There are at least three standards for this measurement. Similarly, which frequency should we use for measuring harmonic distortion? Should we make the measurement at more than one frequency? Which results should we include, and which should we exclude? Standards attempt to answer these questions and allow results to be compared. Engineers love standards – that's why we have so many of them.

If we have designed a piece of equipment, we already know where its likely failings are, so we plan our test to expose those failings. This allows us to measure the failings, make a change to the design, and measure to see if it has made an improvement.

The previous paragraph strikes to the heart of the measurement problem, and raises various points.

- We need to be aware of the limitations of the test equipment. There is no point in attempting to measure the distortion of an amplifier suspected to produce <0.01%THD + N if the test oscillator itself produces 0.01% THD + N.
- We must know the relevance of our measurements. Measuring wow and flutter on an analogue turntable is useful because this measurement exposes known mechanical failings. Measuring wow and flutter on CD players is pointless because they do not suffer from this problem.
- A designer seeking to improve their design makes the test critical. Conversely, the marketing department requests engineering tests that the device is known to pass comfortably – such as harmonic distortion at full output for a CD player, because it gives very good figures.
- Hopefully, nobody understands a given design as well as the designer – who is best placed to decide which tests should be made.
- *Measurements are of most use to the designer.*

For these reasons, measurements quoted by manufacturers or reviewers are not necessarily particularly useful – and this is part of the reason for subjective reviews. (Another reason is that good test equipment is expensive.)

However, if we intend to design and build valve amplifiers, then carefully chosen measurements taken with cheap test gear and carefully interpreted can be very useful indeed.

Choosing the measurement

Transistor amplifiers typically have plenty of global negative feedback to reduce distortion. Because applying feedback can easily turn an amplifier into an oscillator, the amplifier is deliberately made to have an amplitude response that falls with frequency before the feedback is applied. Since negative feedback reduces both linear and non-linear distortions, when it is applied, the frequency response reverts to flatness and non-linear distortion is also reduced. However, because the amplifier's response was falling with frequency before the feedback was applied, less negative feedback is available at high frequencies to correct non-linear distortion. This means that feedback amplifiers *must* have THD that rises with frequency, so a single frequency measurement is inappropriate, and a swept measurement is better.

If we test a circuit that does not have global negative feedback, then a single frequency measurement can be appropriate – if we know what causes the distortion.

A valve distorts because of the curvature in its I_a vs V_{gk} transfer characteristic, and does so at all audio frequencies without fear or favour. Harmonic distortion is the easiest measurement, and we are at liberty to choose any test frequency that we feel is convenient. We might choose 50 Hz or 60 Hz because we have a DVM specified to be accurate to 0.1 dB at that frequency. If we do, we will find that we cannot even measure the fundamental amplitude accurately because stray hum picked up from nearby power wiring beats with our wanted signal to give a gently fluctuating measurement. We need to change our test frequency so that it is clear of mains hum and its harmonics.

Perhaps we could use 10 kHz. This is nicely clear of mains hum, but has problems of its own. Some non-linearities produce mainly higher harmonics, but if the amplifier's amplitude versus frequency response was already falling, this would attenuate the very harmonics we were trying to measure, and give a falsely good result. We need a lower frequency.

In terms of octaves, 1 kHz is in the middle of the audio band, so it is least affected by errors caused by reduced bandwidth and is sufficiently far away from AC mains frequency for hum not to upset the result. Marketing love 1 kHz because it measures very well.

Refining harmonic distortion measurement

Classical harmonic distortion measurements were made at 1 kHz by removing the 1 kHz fundamental and measuring the amplitude of the remaining signal. Although they were appropriate for the valve amplifiers of the time, these tests were rightly criticized when applied to transistor amplifiers because they took no account of the distribution of harmonics and their subjective annoyance.

Weighting of harmonics

Various proposals have been made for weighting the levels of individual harmonics to allow harmonic powers to be summed to give a single figure measure of subjective distortion.

Shorter[1] suggested in 1950 that levels should be weighted by a factor of $n^2/4$ (where 'n' is the number of the harmonic).

$$\text{dB} = 20 \log \frac{V_1}{V_2} = 20 \log \frac{n^2}{4} = 40 \log \frac{n}{4}$$

From n to $2n$ is one octave, so the gradient in dB/octave is:

$$= 40\left(\log\frac{2n}{4} - \log\frac{n}{4}\right)$$

$$= 40\log\left(\frac{\frac{2n}{4}}{\frac{n}{4}}\right) = 40\log 2$$

$$= 12\ \text{dB/octave}$$

Thus, rather than measuring amplitudes of individual harmonics and calculating THD, a rising response of 12 dB/octave could be applied to a conventional distortion meter. In order that the measurement should be comparable with a conventional meter measuring pure 2nd harmonic distortion from a 1 kHz source, the filter would need 0 dB gain at 2 kHz. Note that the combination of the filter and its gain offset means that its weighted distortion measurements are only valid for its specified fundamental frequency.

However, there are problems with the $n^2/4$ weighting technique. Using the 1 kHz measurement example, 20 kHz is a decade higher than 2 kHz, so the filter would add a minimum of 40 dB of gain to harmonics that are inaudible. Since the whole point of the exercise was that the measured result should agree with the subjective nuisance, a 20 kHz low-pass filter is also required.

Although the Shorter recommendation successfully ranked measured distortion against the subjective nuisance, the test suffered from the limitations of its time. The levels of deliberate distortion were quite high (0.41% to 3.7% RMS unweighted) and the loudspeaker was a 'wide-range coaxial-horn' of unspecified distortion.

Peter Skirrow of Lindos Electronics argues that distortion should be measured at 1 kHz using a weighting filter conforming to CCIR468-2 because the response of this filter was determined by the subjective nuisance of different frequencies. Broadly speaking, CCIR468-2 rises with frequency at 6 dB/octave, has 0 dB gain at 1 kHz, and peaks by 12 dB at 6.3 kHz, after which it falls swiftly. See Fig. 3.2.

Summation and rectifiers

The signal that remains after the fundamental has been removed is known as the *distortion residual*, and is composed of a number of harmonically related frequencies. How should we measure the amplitude of this residual? This is not nearly as easy a question as first appears. Perhaps we could measure V_{pk-pk}? See Fig. 3.3.

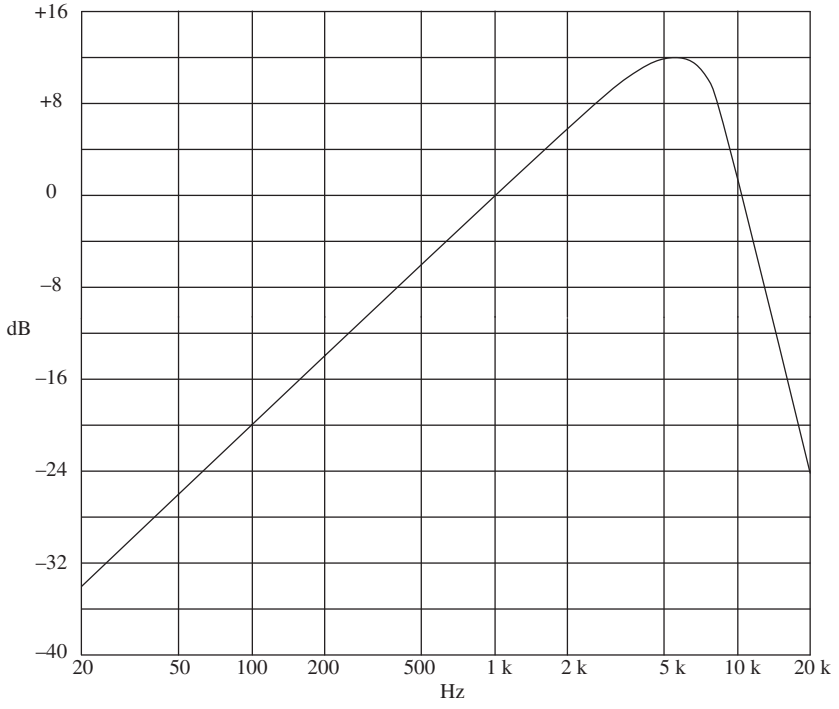


Fig. 3.2 *Frequency response of CCIR468-2 weighting filter*

Both waveforms are square waves with correct harmonics up to the 7th harmonic, and none thereafter, but one has had the phase of the fundamental shifted by 90°, which significantly changes V_{pk-pk}. Mathematically, the correct way to sum individual harmonics is to turn the voltages into powers

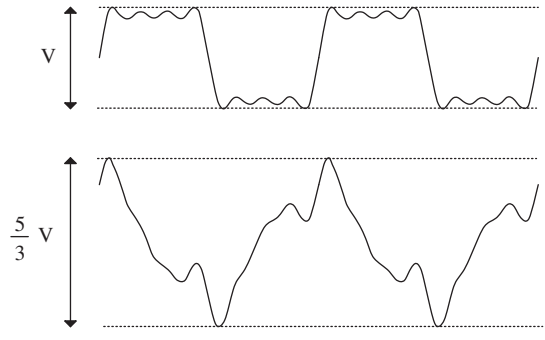


Fig. 3.3 *The effect of phase on waveform shape*

by squaring (V^2/R), take the mean of the powers, then convert this back into a voltage, which implies an RMS measurement. Thus, traditional distortion measurements are made with a meter incorporating an expensive true RMS rectifier, and their measurements reflect this fact by quoting THD in $\%_{RMS}$.

Alternative rectifiers

CCIR468-2 specifies that the rectifier should be peak detecting, because noise is impulsive, and we want to capture the amplitude of these noise spikes. Crossover distortion produces narrow spikes that would contribute very little to an RMS summation, but are subjectively extremely annoying, so the peak detecting rectifier of CCIR468-2 would be ideal for detecting these spikes.

CCIR468-2 is not quite ideal because it needs the previously mentioned gain offset, so the CCIR/ARM recommendation drops the gain of CCIR468-2 by 6 dB to give 0 dB gain at 2 kHz, allowing it to be used for 1 kHz distortion weighting. Most modern test sets allow individual selection of different rectifiers, weightings and offsets, but selecting 'CCIR/ARM' forces all selections to be correct for a subjectively weighted measurement of 1 kHz distortion.

Noise and THD + N

Although the CCIR/ARM method of measuring distortion is cheap and effective, it does have a problem. Properly designed circuitry creates very little distortion. To put it another way, the distortion could be of comparable amplitude to the noise that all electronics generates. When we make our THD measurement, using our meter, how do we know that we are not actually measuring the amplitude of the noise?

One possibility is to view the distortion residual on an oscilloscope. If the waveform appears clean, we are measuring mostly distortion, if a repetitive waveform is difficult to discern, we are measuring noise. Thus, all practical measurements made by a meter are actually THD + N (Total Harmonic Distortion + Noise), and we have to be certain that the noise is sufficiently small to be ignored.

By definition, white noise has constant amplitude with frequency, whereas distortion harmonics occur at very specific frequencies. Our meter is a broadband device, which means that it is sensitive to all frequencies across the audio bandwidth. Thus, although the noise power in a particular frequency band could be quite low, and possibly significantly less than the amplitude of an adjacent distortion harmonic, when summed, the noise powers could easily

swamp the distortion powers. This wouldn't be a problem if it were not for the fact that the ear/brain combination can pick distortion harmonics out of the broadband noise because it works like a *spectrum analyser*.

Spectrum analysers

A spectrum analyser plots amplitudes against frequency, allowing us clearly to distinguish between noise and distortion harmonics. We can now measure individual amplitudes of distortion harmonics, and then apply a subjective weighting such as CCIR/ARM to those numbers mathematically without incurring noise problems.

Analogue audio spectrum analysers were traditionally expensive, but the digital alternative simply relies on raw computing power, and now that this is cheap, many digital oscilloscopes offer options that convert them into spectrum analysers. Alternatively, a PC with a good quality soundcard only needs appropriate software to be able to perform the entire function of distortion measurement and analysis at audio frequencies. Many modern audio test sets are simply outstandingly good soundcards having optimum analogue scaling and dedicated audio analysis software.

However, the process of analogue to digital conversion and its subsequent analysis is *not* transparent, so we need to understand its limitations.

Digital concepts

An analogue signal is continuously variable both in voltage (or current, distance, etc.) and time. Conversely, a digital signal can only change its parameter in discrete steps (*quanta*), and at fixed intervals. Taking measurements to plot a graph is a crude form of analogue to digital conversion, because we freeze the variation, make a numerical measurement, then move on to make another measurement. The power of the technique is that reducing the measurements to a series of numbers allows us to analyse those numbers using a supremely powerful tool – *mathematics* – to find patterns.

Analogue to digital conversion is a two-part process. We freeze the parameter at fixed intervals, and we take numerical measurements of the parameter. These processes can be done in either order. We could take continuous measurements, but only record those measurements that occur at fixed intervals. Alternatively, we can first freeze the parameter at fixed intervals, then make the numerical measurement. It does not matter which way round these two quite distinct processes are applied.

Sampling

The process of freezing the parameter at regular intervals in time is known as *sampling*. If we take 96 000 samples in a second, then the *sample rate* is 96 kS/s, alternatively we can quote the *sampling frequency* as 96 kHz. The sample frequency is significant because the *Nyquist* criterion states that *alias* (fictitious) frequencies will appear if we attempt to sample a signal containing frequencies at, or above, half the sample frequency.

Mild abuse of the Nyquist criterion produces low frequency aliases that were not in the original waveform. You can demonstrate aliasing to yourself by laying two pieces of fine netting one on top of the other and then sliding one against the other. Large circles appear, which are known as *Moiré* patterns (after a type of lace). The reason that Moiré occurs is that one piece of netting is sampling the other, but the sample frequency is the same as the sampled frequency. As the netting slides, the relative phase changes, which changes the frequency of the alias frequencies.

To avoid aliasing, the analogue to digital converter must be preceded by a low-pass filter known, predictably, as an anti-aliasing filter. As an example, a computer soundcard operating at a sample rate of 44.1 kHz must be preceded by an anti-aliasing filter having a cut-off frequency of $\approx$20 kHz. Thus, if we use a computer soundcard for distortion measurements, it is blind to frequencies above 20 kHz. Conversely, digitizing oscilloscopes cannot be preceded by anti-aliasing filters (because their sample rate changes over a wide range), so we must choose either a sufficiently high sample rate that we are confident that aliasing will not occur, or we must add an external anti-aliasing filter.

Scaling

When we plot numbers on graph paper, we have to choose a scale that will conveniently fit our numbers to the lines on the paper. As an example, if the graph paper has ten large squares, each composed of ten small squares, and we had a current measurement ranging from 0 to 8 mA, then we would set a scaling of one large square = 1 mA. This may seem obvious, but what if we chose a scale of one large square = 0.1 mA, or even 10 mA? In the first instance, our data would overload the graph paper, and in the second, it would hardly be seen. The purpose of scaling is to match the range of the parameter to the range of our measurement system.

Similarly, when we convert an analogue parameter to a number, we first scale the parameter to be measured, then we can measure it. Incidentally, this is why most DVMs specify their *basic accuracy* on the 0–5 V range. Their

measurement system actually measures from 0 to 5 V, and the range–switch selects attenuators or amplifiers to scale external voltages or currents to fit this system. Practical problems mean that the scaling cannot be perfect, hence increased errors on all ranges bar 0–5 V.

Quantization

If we have correctly scaled the parameter to be measured, the precision by which we make the numerical measurement is determined by the number of *quanta*, or steps, available. The process of comparing the continuously variable parameter against the series of fixed steps and finding the step that is closest is known as *quantization*. The result of quantizing is a number, although it is commonly known as a digital *word* that is a code for the input voltage, so this is sometimes known as Pulse Code Modulation, or *PCM*.

We now have a succession of digital words appearing at regular intervals that we write into digital memory known as a *waveform record*.

Number systems

Computers count in the *binary* (0, 1) system, rather than the *denary* (0–9) system used by humans. This seems rather limited, because it means that we can count to nine, but no higher, and the computer can only count to one. The solution in both cases is to scale the counting system. Each time we reach 9, and want to add 1, we record the new number as a scaled 1, but this is an inconvenient term, so we call it 'ten'. There is no reason why we should not scale tens, so you will probably remember 'hundreds, tens, and units' from being taught addition in your primary school days. The scaling is shown more formally in the table.

thousands	**hundreds**	**tens**	**units**	**tenths**	**hundredths**	**thousandths**
1000	100	10	1	1/10	1/100	1/1000
10^3	10^2	10^1	10^0	10^{-1}	10^{-2}	10^{-3}

The terms 'hundreds, tenths', etc. are simply powers of the *base*, in this case 10. The binary system works in exactly the same way, but because it uses 2 as its base, rather than 10, its table is slightly different.

32	16	8	4	2	1	1/2	1/4	1/8	1/16	1/32
2^5	2^4	2^3	2^2	2^1	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}

Thus, even though the binary system only counts from 0 to 1, if we use a word with enough *bits*, we can have any number we like.

Precision

Computers use binary numbers, so if we make our numerical measurement more precise by using smaller quantizing levels, there must be more of them, and a binary word comprising more bits is required. CD uses a 16-bit word, and because there are two possible states for each bit, the total number of different levels that can be described by a 16-bit word is $2^{16} = 65\,536$. Similarly, a 24-bit system can describe $2^{24} = 16\,777\,216$ different levels, but requires one and half times as much memory to store each word $(24/16 = 1\frac{1}{2})$.

As a rule of thumb (ignoring dither), the dynamic range (DR) of a digital system is:

$$DR_{dB} = 6n$$

where: n = number of bits.

Thus, a 16-bit system has a theoretical dynamic range of $6 \times 16 = 96$ dB.

We could decide to be more precise by making *more* numerical measurements. Sampling twice as often produces twice as much data.

To sum up, a more accurate description generates a longer waveform record, and this will become significant later.

The Fast Fourier Transform (FFT)

The reason for converting our analogue signal to a digital signal was to allow mathematical techniques to be applied to the resulting numbers and allow patterns to be seen. (Humans are good at recognizing patterns, so any technique that reveals patterns helps understanding.) An oscilloscope allows us to spot patterns that repeat in time, such as a spike that occurs each time a sine wave changes polarity, but a spectrum analyser allows us to spot

patterns in frequency, perhaps a small spike that indicates 5th harmonic distortion.

The FFT is a mathematical tool that converts data initially presented as a graph of a parameter plotted against time into that parameter plotted against frequency. The FFT is immensely powerful, but it has its limitations.

The periodicity assumption

In converting from time to the frequency domain, the mathematics of the FFT make the assumption that the waveform to be analysed repeats itself periodically. This assumption may seem trivial, but it has *major* repercussions.

If we captured a *single* cycle of the waveform, and drew it around a circular drum (like a seismograph) so that the end of the cycle just met the beginning, then by rotating the drum we could replay the waveform *ad finitum* and reproduce our original signal. Unfortunately, any uncertainty as to the precise timing of the end of the cycle will cause a step in level when we attempt to loop the recorded cycle back to itself on replay. However, if we capture more cycles on our drum, the glitch occurs proportionately less frequently and causes less of an error. Thus, capturing a thousand cycles reduces the error by a factor of a thousand, but multiplies the length of the waveform record by the same amount.

Windowing

Another way of reducing the step is to force periodicity by applying a *window* to the waveform record. In this context, a window is a variable weighting factor that multiplies the values of the samples at the ends of the waveform record by zero, but applies a greater weighting (≤ 1) to samples towards the middle. Since any number multiplied by zero is zero, this forces the end samples to zero, and allows the waveform record to be repeated without glitches. See Fig. 3.4.

Because windowing distorts the waveform record, it must distort the results of the FFT. Windowing either spills energy from high amplitude bins into adjacent bins, which produces visible skirts around frequencies having high amplitude, or it changes bin amplitudes. (Because the process of sampling broke time into discrete slices, the results of an FFT must produce frequencies in discrete slices, and these are known as *bins*.) All windows are therefore a compromise between frequency and amplitude resolution.

A window that does not modify sample values is known as a rectangular window (because it multiplies by a constant value of 1 over the entire

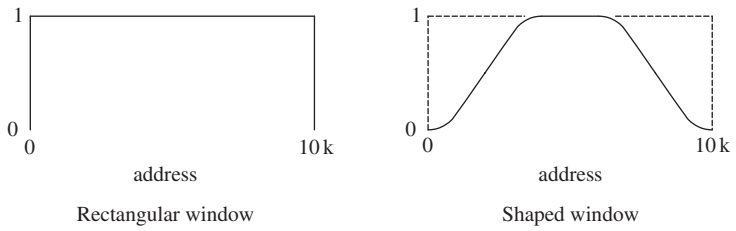


Fig. 3.4 *Windowing forces periodicity*

waveform record). Because the rectangular window does not modify sample values, it does not cause spreading between bins, and it offers the best frequency resolution. Unfortunately, amplitudes are likely to be in error because of periodicity violation. Conversely, the Blackman–Harris window modifies the ends of the waveform record to avoid periodicity violation, which causes spreading between bins, but improves amplitude resolution.

Optimum results are obtained by synchronizing the oscillator to the FFT system so that only complete cycles without phase errors can be captured in the record, allowing a rectangular window to be used. If true *synchronous FFT* is not possible, then a useful compromise is to trigger the analyser from the fundamental frequency and finely adjust oscillator frequency for minimal skirts on the highest amplitude bin.

If multiple waveform records are captured, they can be averaged together to reduce errors. This is a very powerful technique, although it slows measurement speed.

How the author's distortion measurements were made

An MJS401D analogue audio test set and a Tektronix TDS3032 oscilloscope with FFT option were used in combination.

Distortion measurements were made at 1 kHz, and a 400 Hz high-pass filter was engaged to reject hum. The meter used an RMS rectifier to sum harmonic amplitudes correctly, and its bandwidth was restricted to the audible range by a 22 Hz–22 kHz filter. The distortion residual was then passed to the spectrum analyser.

The 9-bit oscilloscope/spectrum analyser used a sample rate of 50 kS/s to maximize the number of cycles captured, and the 22 kHz filter in the MJS401D formed the anti-aliasing filter. It was triggered from the 1 kHz

fundamental, and rectangular windowing was used, with the MJS401D oscillator tuned for minimum skirts to give quasi-synchronous FFT. To clarify results further, the FFTs were averaged over 16 records, each 10 kbits long, resulting in >50 dB of reliable spectrum analyser dynamic range.

Because the dynamic range of the audio test set is added to that of the spectrum analyser, the main limitation is the distortion residual of the test set, so figures below −90 dB should be viewed with caution.

Now that the distortion measurement method is understood, we can use it as necessary to test and compare low distortion valve circuits.

Designing for low distortion

There are many ways of reducing distortion, or to put it less charitably, it's easy to generate distortion inadvertently. To simplify investigation, we will consider the distortion generated by a single stage before progressing to multiple stages. We will consider:

- AC operating point.
- DC operating point.
- Distortion reduction by parameter restriction.
- Distortion reduction by cancellation.
- DC bias problems.
- Individual valve choice.
- Coupling from one stage to the next.

We will investigate each aspect in turn and test our hypotheses with practical measurements.

AC operating point

In theory, the distortion generated by triodes is predominately 2nd harmonic. A common cathode amplifier using 417A/5842 was set up to test the theory. See Fig. 3.5.

Twenty-two 417A/5842 were tested at an output level of +18 dBu (6.16 V_{RMS}), their results were averaged and presented in the table.

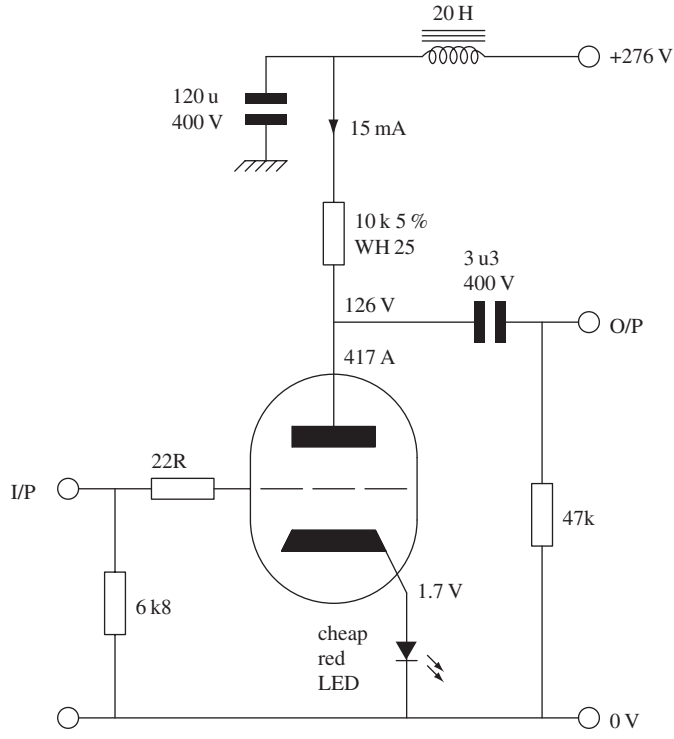


Fig. 3.5 *Common cathode test amplifier circuit*

Harmonic	Level
1st (fundamental)	0 dB
2nd	−41 dB
3rd	−100 dB
4th	−95 dB

The distortion generated by the 417A/5842 clearly *is* dominated by the 2nd harmonic. The 417A/5842 type turned out to be a particularly good example, but even for the worst valves the 2nd harmonic is more than 20 dB higher than any other harmonic. This is useful because it means that we can

reliably use the following formula to predict distortion when drawing and comparing loadlines:

$$\%D_{2\text{nd harmonic}} \approx \frac{V_{\text{quiescent}} - \frac{V_{\text{max.}} + V_{\text{min.}}}{2}}{V_{\text{max.}} - V_{\text{min.}}} \times 100\%$$

The shape of a triode's transfer characteristic is a simple curve ($I_a \propto V_{gk}^{3/2}$), so traversing a smaller distance of the curve is a closer approximation to a straight line, therefore there should be less distortion. This hypothesis was tested by a 7N7/D3a μ-follower circuit. See Fig. 3.6.

In order that the test circuit should not be falsely good when approaching grid current, it was driven from a source resistance of 64 kΩ, thus replicating typical conditions of use. The upper limit of measurement was set by the

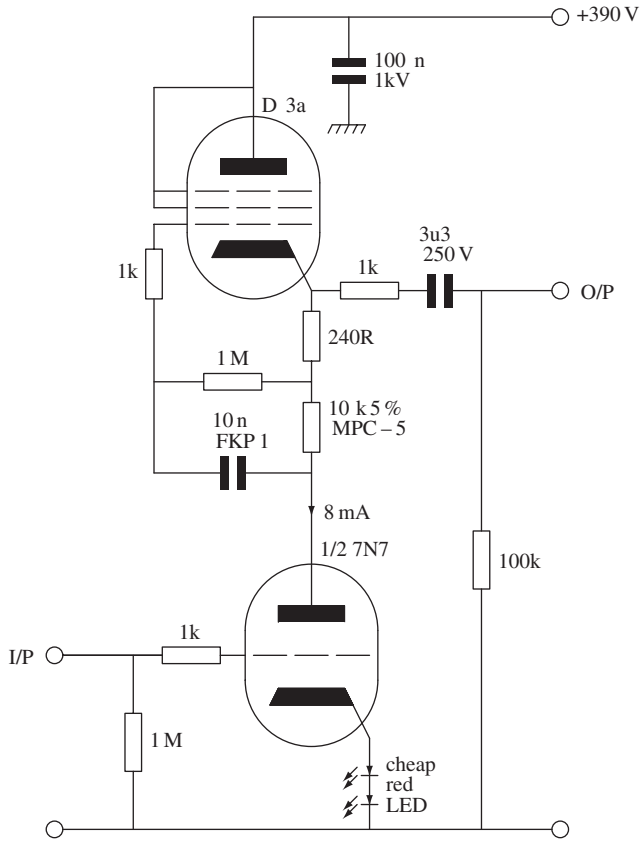


Fig. 3.6 *μ-follower linearity test circuit*

onset of grid current at an output of +34 dBu (THD + N = −43 dB). The lower limit of reliable measurement was set by the ability of the analogue analyser to lock cleanly to the distortion waveform, which began to degrade at an output of +14 dBu (THD + N = −63.5 dB). Between these limits, the output level was changed in 1 dB steps, and a graph of THD + N was plotted against output level. See Fig. 3.7.

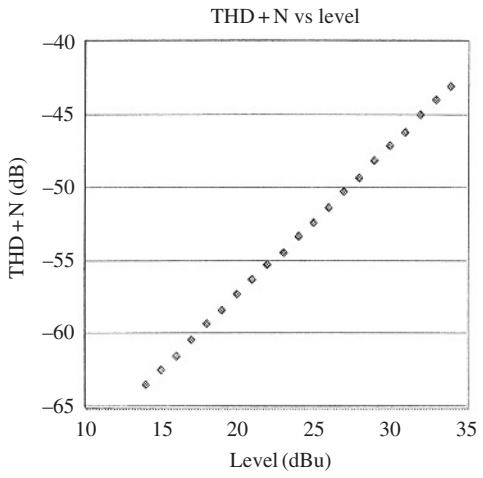


Fig. 3.7 *Graph of distortion vs level for μ-follower test circuit*

The graph clearly shows that THD + N is directly proportional to output level. Thus, a distortion measurement of 1% at 15 V_{RMS} implies distortion of 0.1% at 1.5 V_{RMS}. This fact is extremely useful if it is necessary to estimate the distortion of a triode handling small-signal voltages – such as would normally be encountered early in an RIAA stage.

The supposition that triode distortion is predominately 2nd harmonic and is proportional to level is true for all triodes when used with practical resistive anode loads. The effect of an active load ($R_L \Rightarrow \infty$) is to suppress the 2nd harmonic, but barely change higher harmonics. Once the 2nd harmonic has been suppressed, the effects of the higher harmonics become more significant, with the result that some triodes do *not* then have distortion that is proportional to level. If you use an active load, you may need to check whether distortion remains proportional to level for that particular type of valve.

DC operating point

Distortion changes with V_a and I_a will be investigated later because they are caused by changes in the small-signal parameters μ, r_a and *gm* that are normally assumed to be constant. Thus, unless we also need to maximize voltage swing, our choice of operating point simply needs to be wary of grid current and cut-off. The problems of cut-off are obvious, but grid current causes far more problems.

Distortion due to grid current at contact potential

As V_{gk} approaches 0 V, grid current begins to flow, and the input resistance of the valve falls dramatically. If the source driving the valve had $r_{out} = 0$, this would not be a problem, but it is highly likely that it has significant output resistance, and the potential divider that is momentarily formed at the positive peaks of the waveform where grid current flows clips the input signal. Symmetrical clipping produces a square wave composed of odd harmonics, but grid current clips asymmetrically, so even harmonics can also be expected.

The distortion caused by grid current is obnoxious because it is composed of high order harmonics. The following traces were obtained by driving the lower valve of a μ-follower into grid current from a source resistance of 47 kΩ. The input signal was increased until distortion of the output waveform was just visible on a carefully focused analogue oscilloscope. The THD + N was measured to be 2%, and the distortion residual had a very distinctive waveform. See Fig. 3.8.

When analysed, the distortion residual revealed a spray of high level, high order, odd and even harmonics. See Fig. 3.9.

Although grid current occurs at $V_{gk} = 0$ V in an ideal valve, practical valves enter grid current a little earlier due to the thermocouple effect of junctions between dissimilar heated metals within the valve, and the average energy of electrons in the electron cloud above the cathode surface. Typically, grid current begins at $V_{gk} \approx -1$ V, and this is known as the *contact potential*. Measuring distortion whilst driving from a high impedance source is an excellent way of determining contact potential for a given valve.

Distortion due to grid current and volume controls

It is perfectly possible to change the design of volume control preceding an amplifier stage and *measure* a change in distortion. The most common type of

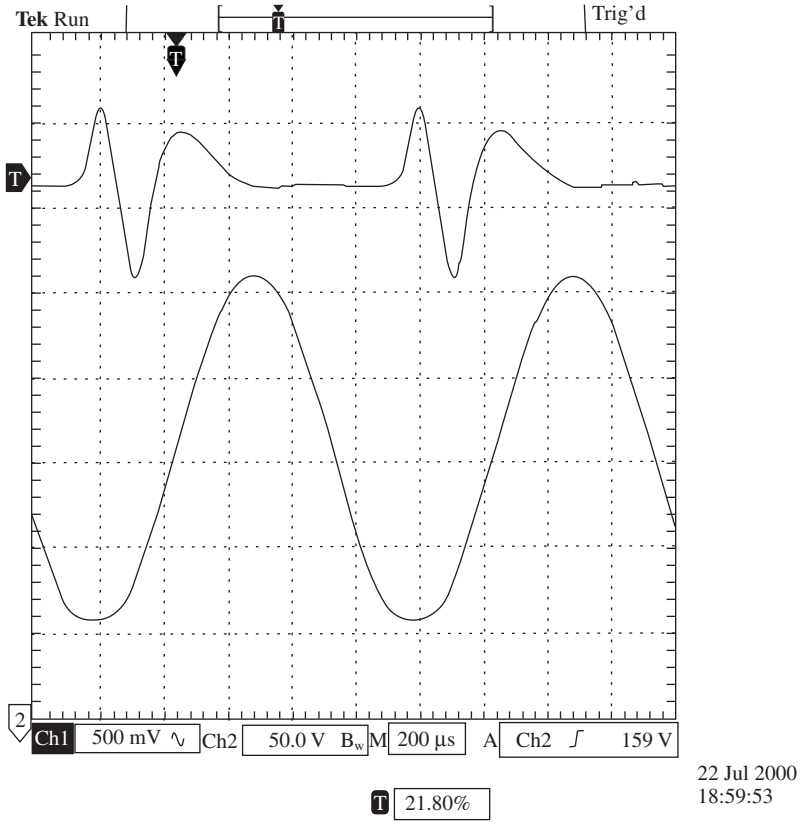


Fig. 3.8 *Upper: the distinctive distortion waveform caused by grid current. Lower: soft clipping caused by grid current*

volume control is a resistor from which a variable tapping is taken, either a wiper moving along the resistive track, or a switch wiper selecting a tapping from a chain of fixed resistors. See Fig. 3.10a.

Alternatively, we can use a fixed series resistor followed by a variable shunt resistor. See Fig. 3.10b.

Unfortunately, the circuit of Fig. 3.10b has much higher output resistance than that of Fig. 3.10a. Measuring distortion whilst driving from a high source resistance is a very sensitive test of gas current because the (non-linear) gas current develops a voltage across the source resistance, which is in series with the signal. As the source resistance rises, so does the distortion.

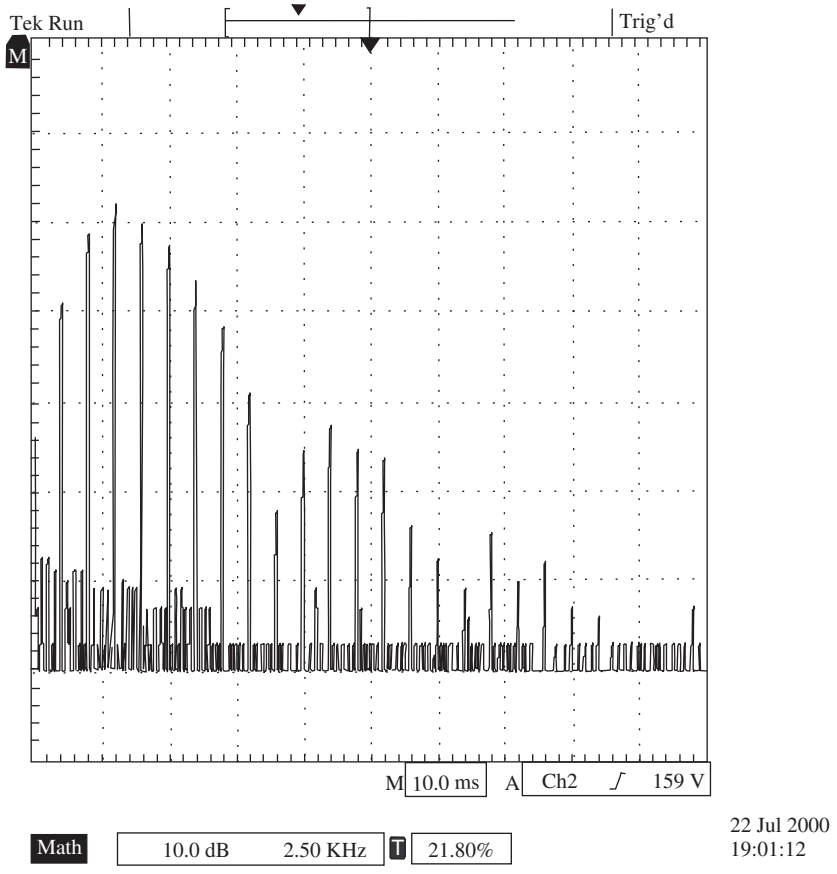


Fig. 3.9 *Distortion spectrum produced by grid current acting on 1 kHz sine wave. Vertical scale: 10 dB/div. Horizontal scale: 2.5 kHz/div. (DC–25 kHz)*

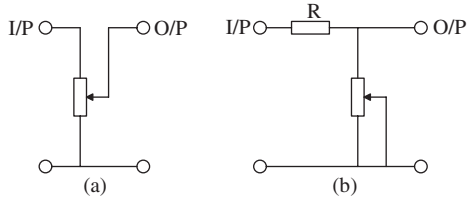


Fig. 3.10 *Fundamental basis of most volume controls*

A 6C45П cathode follower biased by an EF184 constant current sink was tested at +20 dBu (7.75 V_{RMS}). When driven from a 5 Ω source, the distortion was 0.02%. A 100 kΩ type (a) volume control has a maximum output resistance of 25 kΩ, so distortion was also measured with a 25 kΩ source resistance, and found to be unchanged. However, when the source resistance was increased to 1 MΩ, the distortion rose to 0.2%. Admittedly, it is unlikely that the source resistance would be as high as 1 MΩ, but 100 kΩ would be quite possible from the volume control in Fig. 3.10b.

Operating with grid current (Class A2)

Most Class A amplifiers operate without grid current because this gives a high grid resistance that is easily driven. Once V_{gk} becomes positive, rather than repelling electrons from the cathode, the control grid becomes an anode and attracts electrons, most of which are then captured by the true anode that is at a much higher voltage, but some flow out of the grid as grid current. Grid current has important consequences:

- The electron stream from the cathode is divided between grid and anode current, implying partition noise. However, the most likely use of Class A2 is in the output stage, where signal voltages are high, so this noise is unlikely to be a problem.
- There is a potential difference between the grid and the cathode (V_{gk}), and current flowing through the grid (I_g), so it must be dissipating power in exactly the same way as an anode. If the grid was not designed to dissipate power, it will quickly heat, distorting its shape and possibly destroying the valve.
- Because the input resistance of the grid is very low, imposing a signal voltage on the grid requires considerable power ($P = V^2/R$), which must be provided by the driver stage.
- However, because the grid is driven positive, it is possible to drive the anode of a triode far closer to 0 V than if V_{gk} was negative. The efficiency of the output stage is thus significantly increased.

Driver stages for Class A1 are voltage amplifiers that only need to supply sufficient current to charge and discharge the Miller capacitance of the output stage, but a driver stage for Class A2 must provide power. There are two ways in which this power can be delivered.

The driver stage can be designed to be a small power stage. One possible choice is the common cathode dual triode 6N7 that can either be operated in push–pull, or single ended with the two triodes paralleled to double P_a.

A transformer reflects its load impedance by n^2, so a step-down transformer with a voltage ratio of 2:1 will increase the load impedance seen by the driver valve by a factor of 4. Because a transformer in the anode circuit of a valve theoretically allows V_a to swing to $2\,V_{HT}$, requiring double the anode swing is not a problem. Additionally, the low R_{DC} of the secondary reduces the chances of thermal runaway in the output stage. Sadly, good driver transformers are even more difficult to design than output transformers because they operate at higher impedances.

Alternatively, the Class A2 stage can be driven DC coupled from a cathode follower. A power valve is still required, but it no longer needs to be able to swing many volts. Power frame-grid valves that have high mutual conductance, but low $V_{a(max.)}$, such as the 6C45П and E55L, are ideal as power cathode followers. Unfortunately, frame-grid valves tend to have modern, efficient heaters, which means that their $V_{hk(max.)}$ is quite low, possibly causing a problem if the Class A2 stage requires significant grid voltage swing. To bias the Class A2 stage correctly, the cathode of the cathode follower can only be

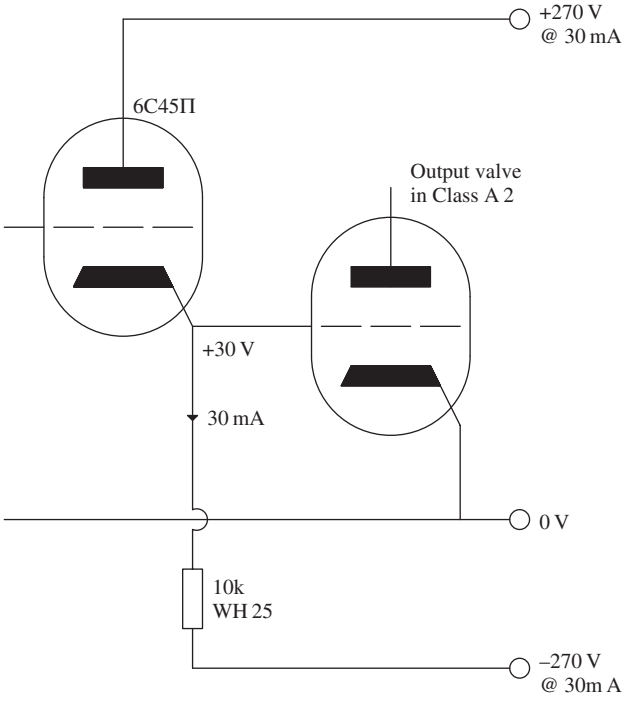


Fig. 3.11 *Using a DC coupled power cathode follower to drive Class A2*

slightly positive, but we need a reasonably large value of R_L to ensure linearity of the cathode follower, so a negative supply is required. See Fig. 3.11.

Both of the previous solutions offer a low output resistance, but it is not zero. Because $r_{out} \neq 0$, it forms a potential divider with the input resistance of the Class A2 stage, causing attenuation. If V_{gk} swings negative, the input impedance of the Class A2 stage becomes infinite, and there is no longer any attenuation, causing distortion. There is no advantage to using a constant current sink load for the cathode follower because it faces the low resistance load of the Class A2 grid. Whereas a Class A1 stage should never be driven into grid current for fear of distortion, the Class A2 stage must never be allowed to stray *out* of grid current, or distortion will result.

Distortion reduction by parameter restriction

Triodes produce primarily 2nd harmonic distortion because as r_a changes with I_a, the attenuation of the potential divider formed by r_a and R_L changes, with more attenuation on one half cycle of the waveform than on the other. However, there are ways of reducing this distortion:

- Use a large value of R_L. If $R_L >> r_a$, then the changing attenuation of the potential divider is insignificant because the attenuation itself becomes negligible.
- Hold I_a constant so that r_a cannot vary. This implies an active load such as a constant current source, and is the basis of the μ-follower.

These two methods are actually very similar because both seek to make $R_L >> r_a$. (For an ideal constant current source, $r_{internal} = \infty$.) In general, for a given HT voltage and I_a, a constant current source made with a single valve can be expected to reduce distortion by a factor of ≈ 7.

Once the previous methods of distortion reduction have been used, the amplifying valve sees an almost horizontal AC loadline, and when $R_L > 50r_a$, the far lesser effect of variation of μ with V_a becomes observable. The variation of μ with V_a can be reduced by avoiding operation at low I_a (where the anode curves begin bunching), and by choosing a valve whose curves bunch less as I_a tends to 0. See Fig. 3.12.

Bunching of anode curves is caused by the inevitable non-uniformity of the electric field between the grid wires at the grid/cathode region, so the graph compares two GEC directly heated triodes having similar μ, but the solid curves are due to a grid wound with a few turns of coarse wire, and the dashed curves due to a grid wound with more turns of fine wire.

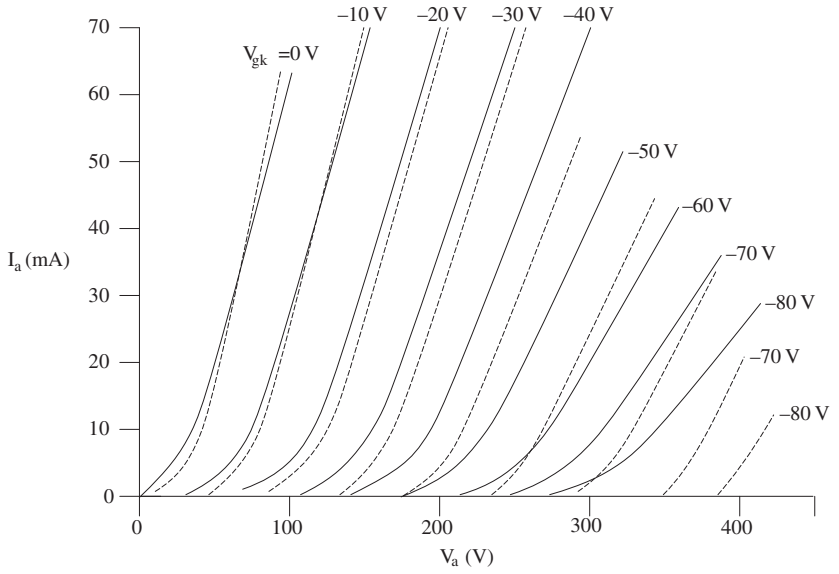


Fig. 3.12 *Grid construction: more fine turns (dashed) vs fewer coarse turns (solid).* (After Henderson[2] GEC)

Unfortunately, as the grid wire becomes finer, it is less able to support itself, but a frame grid allows arbitrary thickness of wire, which is why valves such as the E88CC, and particularly the 6C45Π, exhibit very little bunching.

Alternatively, it may be possible to hold V_a constant. Clearly, this cannot be done if the stage has gain, but a cathode follower can be arranged to have constant I_a and V_a simultaneously.[3] See Fig. 3.13.

The middle valve is the cathode follower. The lower valve is the traditional pentode constant current sink that forces I_a in the cathode follower to be constant. The upper valve is also a cathode follower, and should have high μ and *gm*, so the 6C45Π (μ = 52) is ideal. The upper valve sees a high impedance load, so its gain is:

$$A_v = \frac{\mu}{(\mu + 1)} = \frac{52}{(52 + 1)} = 0.98$$

The upper cathode follower's grid is AC coupled to the output of the middle cathode follower, and because its gain is almost unity, its cathode is at the

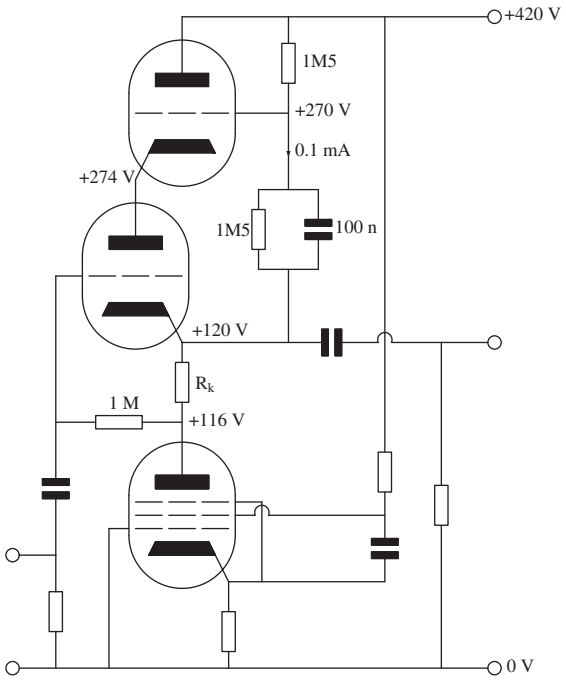


Fig. 3.13 *The cathode follower allows constant I_a and V_a to be forced*

same AC voltage as its grid. Thus, even when the middle cathode follower swings its cathode, the upper cathode follower forces its anode to swing by an almost identical amount, and constant V_a has been enforced simultaneously with constant I_a.

Unfortunately, the improvement is accompanied by significant costs:

- The required HT voltage has been raised by V_a of the upper cathode follower.
- We need a third elevated heater supply (for the upper cathode follower).
- Cathode followers are already prone to instability, and bootstrapping the anode of one with the output of another invites further problems.

You might have a different opinion, but the author feels that a carefully designed cathode follower sitting on a pentode already challenges his test equipment.

Distortion reduction by cancellation

In theory, if two common cathode triode amplifiers are operated in a cascade, because each stage inverts, the distortion of the second triode is inverted with respect to that produced by the first triode, and cancellation should occur. However, a moment's thought shows that this is unlikely to occur to any significant extent. Distortion is proportional to level, and because the second triode has gain, it produces a significantly higher level, and therefore proportionately higher distortion than the first triode. A small amount of cancellation may occur, but the improvement is $\propto 1/A_2$, thus, if the second triode was a type 76 ($\mu = 13$), and $A_v = 10$, we might reduce distortion from 1% to 0.9%, which is probably less than the sample to sample variation of distortion in either valve.

Perhaps we could choose the second valve to be much more linear than the first, so that they produce equal amounts of distortion? Low μ valves are the most linear, so an 845 ($\mu = 5.3$) might achieve $A_v = 4$, so we need a valve that produces four times the distortion of the 845. This can probably be done, and adjusting the bias of the first valve would allow complete cancellation to be achieved. However, this cancellation would be critically dependent on the gain of the 845, which is determined by R_L, but R_L is a loudspeaker whose impedance changes with frequency. In practice, 6 dB reduction in 2nd harmonic distortion is feasible.

Distortion cancellation is only achieved reliably if the two valves are identical, carry exactly the same signal, and have the same load conditions.

Push–pull distortion cancellation

The push–pull transformer-coupled Class A output stage meets most of the requirements for distortion cancellation. However, unless the two valves are an accurately gain-matched pair, or provision has been made for DC and AC balance, cancellation will not be perfect. Nevertheless, 14 dB cancellation is routinely achieved because the tight coupling between the two halves of the transformer assists AC balance.

Differential pair distortion cancellation

The differential pair with constant current sink tail provides optimum conditions for distortion cancellation because the signal *current* is forced to swing between the two valves with no loss. Provided that the load impedances are

matched, the voltage swings at each anode must be equal and opposite, theoretically allowing perfect cancellation of 2nd harmonic distortion. The anode load resistors can easily be matched to 0.2% by a DVM, and if each anode drives a cathode follower, then the shunt capacitance is so small that any imbalance is insignificant at audio frequencies. (Even at 20 kHz, $X_c = 1.6\,M\Omega$ for the 5 pF input capacitance of a typical cathode follower, so this is significantly larger than the typical 47 kΩ anode load resistors.)

A Mullard 6SN7GT with well-matched sections was compared in different configurations with $I_a = 7.5$ mA, $V_a = 230$ V, and a swing of +14 dBu at the anode. The differential pair was measured between the anodes at a level of +20 dBu, corresponding to +14 dBu at each anode. See Fig. 3.14.

Harmonic	Common cathode	Differential pair	μ-follower
2nd	−51 dB	−77 dB	−68 dB
3rd	−93 dB	−89 dB	–
4th	(−106 dB)	–	–

As can be seen from the table, the differential pair cancelled even harmonic distortion by 26 dB, but sums odd harmonic. Although 0.0035% 3rd harmonic is unlikely to be a problem, it indicates that differential pairs are ideally built with valves that produce small amounts of odd harmonic distortion. Conversely, the μ-follower was not as effective at reducing 2nd harmonic, but all other harmonics were below the limits of reliable measurement.

DC bias problems

Having chosen the topology of a stage with great care, we choose an operating point that cunningly maximizes output swing, minimizes distortion, uses standard component values, and all within the current capability of the power supply. We now need to bias the stage, which can be done in a number of ways:

- Cathode resistor bias.
- Grid bias.
- Cathode bias with a rechargeable battery.

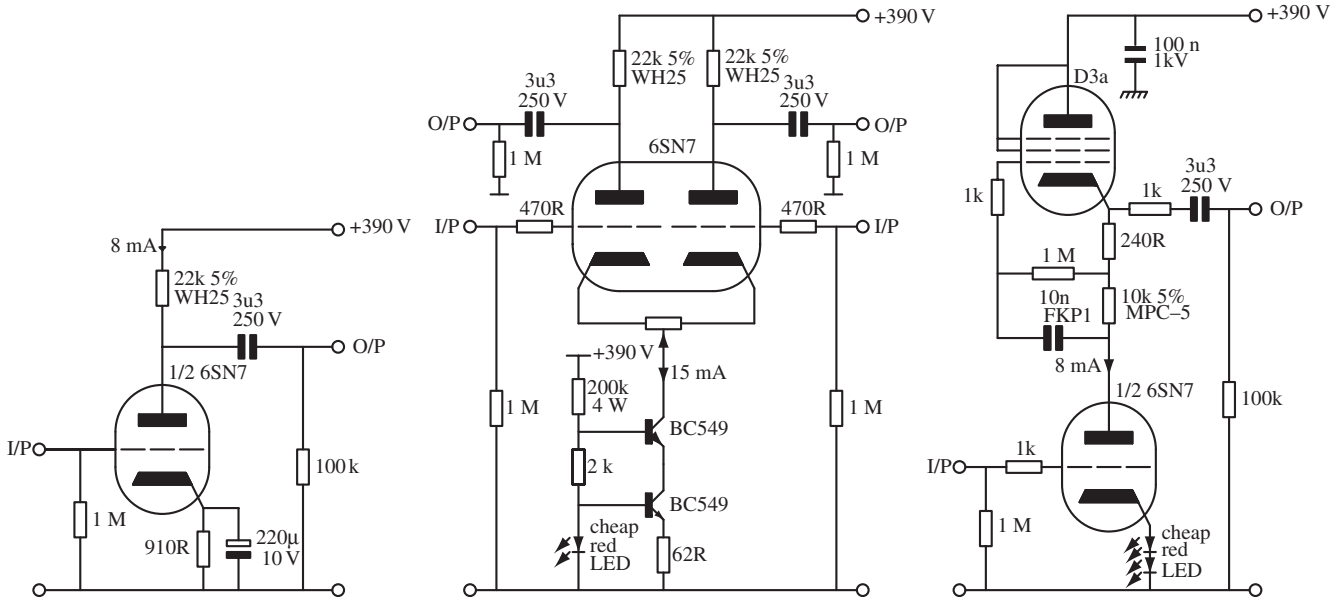


Fig. 3.14 *Test circuits for distortion comparison*

- Cathode bias with a diode.
- Cathode bias with a constant current sink.

Cathode resistor bias

Bias can be achieved by inserting a resistor in the cathode path. See Fig. 3.15.

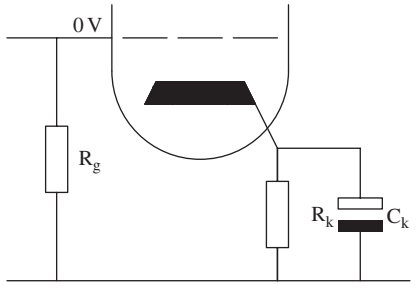


Fig. 3.15 *Cathode bias using a resistor*

If valve current rises, resistor current also rises, making the cathode more positive with respect to the grid, thus tending to turn the valve off, and offering some overcurrent protection. This method of bias has the least sensitivity to variations between valves, making it by far the most popular bias choice. We know I_a and the required V_{gk}, so we simply apply Ohm's law to determine the required cathode resistor.

However, inserting a resistance in the cathode circuit of a single valve common cathode amplifier creates negative feedback that reduces gain, which might not be acceptable. The traditional solution bypasses the resistor with a capacitor (which is a short circuit at audio frequencies), the cathode is connected to ground at AC and negative feedback is prevented. It is generally argued that the audio bandwidth extends from 20 Hz to 20 kHz, and that audio electronics should be as nearly perfect as possible within this bandwidth. The (inevitably) electrolytic cathode bypass capacitor should not be allowed to contribute to any filtering action, so its value is usually set to produce $f_{-3\,dB} = 1$ Hz.

When biasing a stage, we make the assumption that the signal voltage is sufficiently small that it does not affect the DC conditions. However, as clipping is approached, the signal voltage at the anode of a triode could be hundreds of volts peak-to-peak, and the distortion (which contains a DC component) temporarily lowers the average V_a. A secondary effect of this

distortion is that it has a DC component that changes the mean anode current.

As an example, a common cathode triode amplifier was tested. When the generator was muted, $V_a = 117.1\,V$, but when the stage was driven to a level that produced 5% THD + N, the mean anode voltage fell to 114.2 V, and changed the mean anode current. The change in mean anode current is integrated by the cathode capacitor. Since $f_{-3\,dB} \approx 1\,Hz$, this implies $\tau \approx 160\,ms$. When the overload passes, the capacitor takes $5\tau \approx 1\,s$ to recover to 99% of the previous bias point. During this time, r_a (which is dependent on I_a) will have changed, slightly changing r_{out}. If the circuit feeds a passive equalization network, r_{out} is inevitably part of the design, so the change in r_{out} causes a temporary frequency response error. Although a minor frequency response error could be considered irrelevant when the amplifier is producing 5% THD + N, a frequency response error that decays to zero over a period of 1 s after overload might not be so acceptable.

The bias shift effect can be observed by measuring the DC voltage across the cathode bypass capacitor with and without a large signal at the anode. This method has the advantage that an ordinary DVM can be used, whereas measuring at the anode requires a DVM that can measure DC accurately in the presence of significant AC.

Ideally, there should never be a shift in the operating point of a valve, whatever the signal level. Provided that the valve is never driven to produce >1% THD, cathode bias is perfectly satisfactory, but if clipping is likely, an alternative bias strategy should be considered.

Grid bias ($R_k = 0$)

If $R_k = 0$, the DC component of distortion cannot cause bias shift.

Grid bias from an auxiliary low current negative supply is common in the output stages of Class AB power amplifiers, whereas lithium battery bias is occasionally found in pre-amplifiers. See Fig. 3.16.

Grid voltage is fixed, and valve current is determined purely by valve characteristics, so it there is no protection against overcurrent, or compensation for changes in valve characteristics with age.

Overcurrent protection is important in output stages because the winding resistance of an output transformer is negligible and the output valve is almost certainly being operated at maximum anode dissipation. Current from the supply is therefore almost unlimited and a fault is likely to damage an expensive valve quickly.

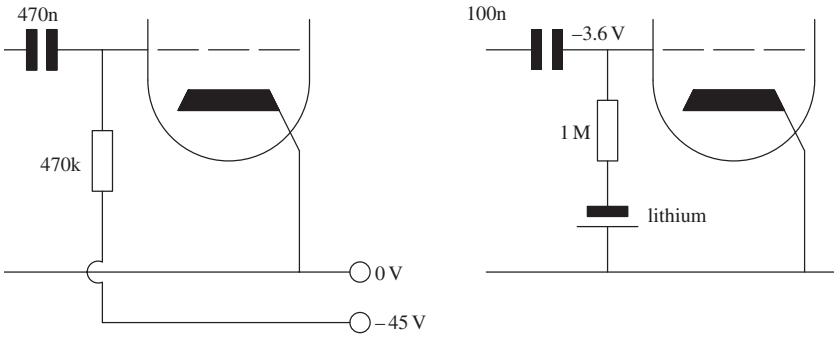


Fig. 3.16 *Grid bias using an auxiliary power supply, or a lithium battery*

Conversely, in pre-amplifiers or driver stages, operating the valve at much more than half maximum anode dissipation is unusual, and the anode load is typically a resistor that significantly limits any fault current. It is quite conceivable that a fault resulting in maximum current could leave the valve operating well within its limits, and no damage at all would occur.

Rechargeable battery cathode bias ($r_k = 0$)

Rechargeable cells have extremely low internal resistance, so if they are inserted in the cathode path, they do not allow a feedback voltage to appear. See Fig. 3.17.

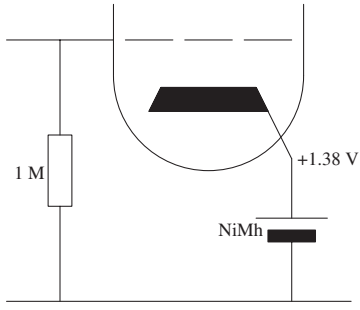


Fig. 3.17 *Cathode bias using a rechargeable battery operated at trickle current*

Although the diagram shows only one cell, a number of (identical) cells could be connected in series to set the required voltage, although this would be rather bulky. Provided that $I_k \leq C/10$ (C is the cell capacity in A/hr), the

self-heating caused by continuous charging will not damage the cell. However, since the cell is in a valve amplifier, it is probably warmer than the battery manufacturer expected, so limiting the current to $C/20$ might be wise. An AA size nickel metal hydride (NiMh) cell develops ≈1.38 V when charged continuously at 15 mA.

Diode cathode bias ($r_k \approx 0$)

Rather than using a resistor, we can use a diode for cathode bias. See Fig. 3.18a.

The advantage is that the diode's internal resistance is much lower than the traditional cathode resistor, so we no longer need to bypass it with a capacitor. Although diode slope resistance is low, sometimes it may be necessary to calculate its effect on r_a. The table shows forward drops and slope resistances (r_{diode}) for various diodes:

Diode type	**Forward drop @ 10 mA**	**Typical r_{diode} @ 10 mA**
Silicon small-signal diode (1N4148)	0.75 V	6.0 Ω
Germanium small-signal diode (OA91)	1.0 V	59 Ω
Infrared LED (950 nm)	1.2 V	5.4 Ω
Cheap red LED	1.7 V	4.3 Ω
Cheap yellow yellow/green LED	2.0 V	10 Ω
EZ81	2.3 V	195 Ω
True green LED (525 nm)	3.6 V	30 Ω
Blue LED (426 nm)	3.7 V	26 Ω
EZ80	5.5 V	485 Ω

The more esoteric LEDs provide a higher forward voltage drop at the expense of a much higher internal resistance, so it is better to use a pair of cheap red LEDs in series than an (expensive) chromatically true green or blue LED. Germanium is unlikely to be useful because for an almost equal voltage drop, an infrared LED has one tenth of the internal resistance. The thermionic diodes barely improve on a pure resistance, so they are unlikely to be useful.

Reverse bias produces more noise in a diode than forward bias, but enables higher reference voltages. See Fig. 3.18b.

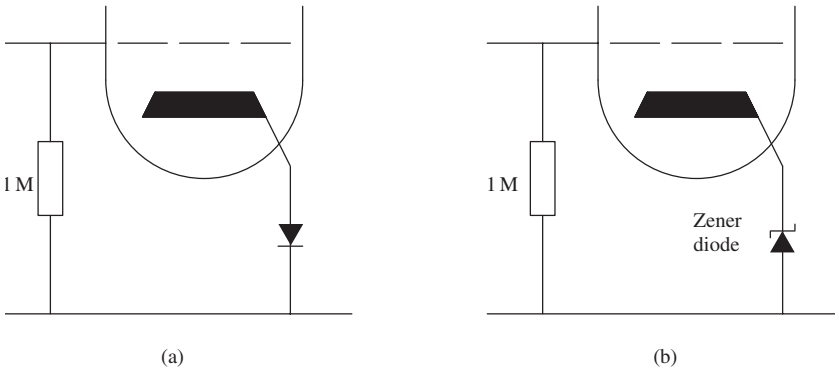


Fig. 3.18 *Cathode bias with a diode*

Low voltage Zener diodes use true Zener action, but higher voltage diodes actually use the avalanche effect. At 6.2 V, both effects are present, their opposing temperature coefficients cancel, r_{diode} is at a minimum, and so is noise, so 6.2 V Zeners are the best. If an ideal higher voltage reference is required, it may be better to have a string of 6.2 V Zeners and LEDs than a single high voltage Zener.

Diode bias is ideal for biasing the lower valve of a μ-follower or SRPP because I_a is stabilized by the bias arrangements of the upper valve.

Because $r_{diode} \neq 0$, a change in signal current causes a change in the voltage across the diode. The signal current also produces the voltage across R_L, so:

$$\frac{v_{r_{diode}}}{r_{diode}} = \frac{v_{R_L}}{R_L} = i_{signal}$$

Cross-multiplying:

$$v_{r_{diode}} = \frac{v_{R_L} \cdot r_{diode}}{R_L}$$

The significance of this equation is that r_{diode} changes with applied current. See Fig. 3.19.

Because r_{diode} is not constant, the signal voltage developed across it is distorted. Unfortunately, this distorted signal voltage is in series with the input signal because the valve amplifies the *difference* in voltage between the grid and cathode. See Fig. 3.20.

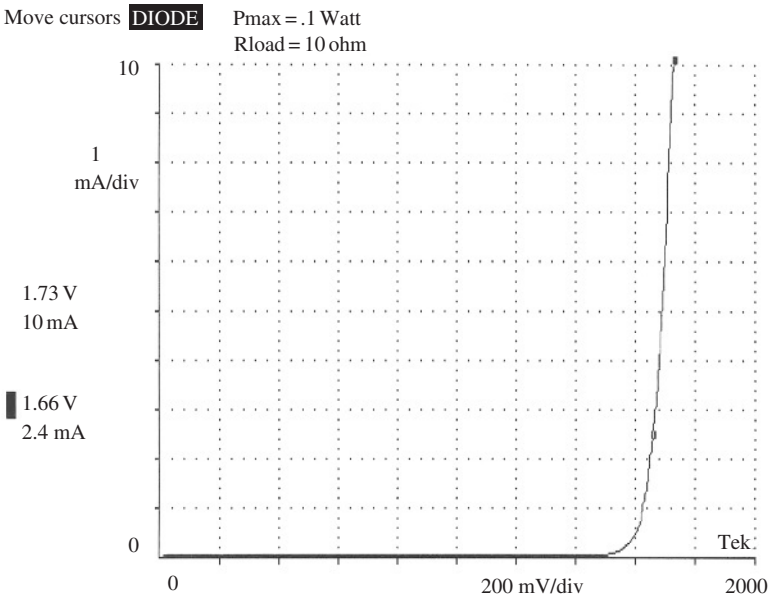


Fig. 3.19 *Current against applied voltage for cheap red LED*

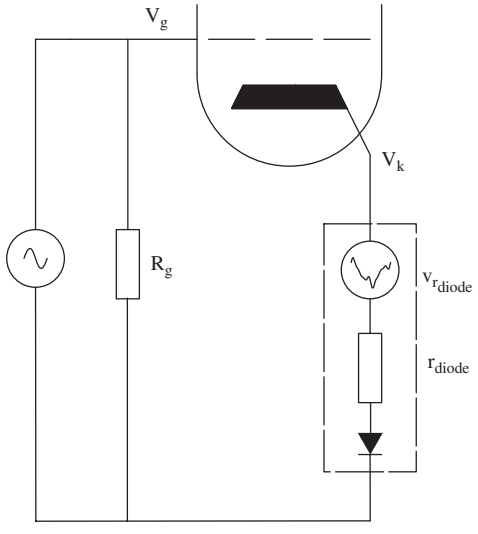


Fig. 3.20 *Non-linear diode internal resistance adds distortion in series with the source*

However, the equation and diode curve show us that the distortion added by the diode can be reduced by:

- Avoiding diode bias for $I_a < 10\,mA$ (because $r_{internal}$ is particularly variable at low currents).
- Maximizing R_L.
- Reducing the output signal voltage v_{R_L}.

These conditions imply that diode bias is best suited to:

- RIAA input stages: I_a is high and signal levels are low. Additionally, the stage can recover instantly from clipping due to high voltages at high frequencies caused by dust, etc. on the record.
- μ-follower stages: the active load maximizes R_L, and I_a is likely to be high.

Constant current sink bias

Cathode current can be forced to the design value despite valve parameters or fault conditions using a constant current sink. See Fig. 3.21.

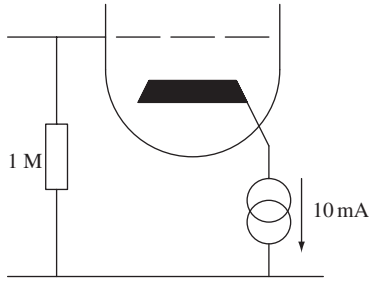


Fig. 3.21 *Cathode bias using a constant current sink*

However, because a constant current sink is an open circuit to AC, it causes 100% negative feedback in a single-ended stage, but it is unsurpassed for biasing a differential pair.

Individual valve choice

Although a set of anode characteristics having noticeably different spacings between the curves indicates distortion, evenly spaced characteristics do not

necessarily imply low distortion. Ultimately, we must either use valves designed for low distortion, or test for distortion.

Which valves were explicitly designed to be low distortion?

Minimizing distortion costs money, so when low distortion valves were designed they were targeted specifically at the audio market, which included the broadcast, recording and film industries, and, of course, the consumer.

In the 1930s, gain was extremely expensive. The idea of deliberately throwing gain away (negative feedback) was treated as heresy, so much so that although Harold Black's jotted notes were witnessed on 18th August 1927, his US patent[4] was not issued until 21 December 1937. As a consequence, low distortion was reliant on valve design and construction, so valves like the 76 were designed to be low distortion. As feedback became more widely accepted, it became cheaper to reduce distortion by sacrificing gain, so the final generation of valves (ECC83/12AX7, etc.) had higher gain, but worse distortion.

Low distortion valves were required by the telecommunications companies, but not because they were concerned with the fidelity of baseband audio. If we need to provide a thousand telephone circuits between two cities ten miles apart, we could lay 1000 twisted pairs, but a cable containing this amount of wire is expensive and cumbersome to lay. The solution adopted by the telecommunications companies was to modulate each telephone circuit onto an RF carrier with its own frequency – just like different radio stations. The 1000 modulated carriers could then be passed down a single (usually co-axial) cable which was cheap and easily laid. All cables introduce loss, and between cities the loss becomes significant, so all cables needed repeater amplifiers at regular distances. One of the many advantages of multiplexing a thousand telephone circuits onto one cable was that only one repeater amplifier was needed instead of a thousand. However, any distortion in that amplifier would cause one telephone conversation to crosstalk onto another. Valves designed for use in broadband telephone repeater amplifiers were therefore required to produce low distortion.

Many of the final generation of valves used a frame grid; some, such as the 417A/5842, were designed for low distortion, others, such as the ECC88/E88CC, simply benefited from improved production engineering and produce usefully low distortion. Some valves such as the E182CC and 6350 were designed for use in digital computers, where the most important consideration was long life even with full heater power and no anode current, which tempts the growth of cathode interface resistance.

Other valves were designed and manufactured with a complete lack of regard for distortion.

The problem of field, or vertical, scanning in televisions is very similar to that of an audio amplifier driving a loudspeaker. Both use a transformer to couple to the driving valve, and the frequency range is similar. However, television scan coils are driven by controlled current, rather than applied voltage as is conventional for loudspeakers. Unfortunately, the finite primary inductance L_p of the output transformer draws a current in addition to the scan coil current, and this means that the total current needed from the field scan valve is distorted compared with the ideal current required by the coils. There were many ways of achieving this distortion, but one was to use the curvature of the I_a/V_a characteristic of a triode. Since transformer L_p was not tightly controlled, the required distortion had to be controllable, so a variable resistor was often inserted in the cathode circuit of the valve to allow adjustment of vertical display linearity.

The crux of the previous argument is that there was no requirement *whatsoever* for the valve manufacturers to produce field scan valves with outstanding, or even consistent linearity, since this had to be individually adjusted for each television. Early field scan valves such as the dual triode 6BX7 show wide variations in distortion (4:1 between worst and best), so they have to be selected for audio use, and the probability of finding a pair of low distortion valves in one envelope is low, so selecting a pair of low distortion 6AH4 single triodes would be a much cheaper alternative. Later generation valves such as the ECC82 (also intended for use as a field scan oscillator) benefited from improved production techniques and distortion is extremely consistent from sample to sample: it is consistently poor.

Carbonizing of envelopes

Deketh[5] pointed out that not all electrons accelerated from the cathode/grid interface strike the anode – some miss and collide with the envelope, causing secondary emission. Secondary emission is important because it means that the envelope acquires a negative charge that distorts the flight of electrons from cathode to anode. Deketh considered distortion at high amplitudes in power valves and showed that carbonizing the inside surface of the envelope was beneficial because it reduced secondary emission. At the time, nobody was worried about audio distortion at <1%, and Deketh might not have had access to an audio spectrum analyser, so he did not publish distortion results at lower levels. Nevertheless, the author's measurements at +28 dBu($\approx$19.5 V_{RMS}) show significantly reduced (≈ -6 dB) distortion

for samples of the 6SN7 having a carbonized envelope compared to clear envelopes.

Deflecting electrons

Amplifying valves control the flight of electrons by imposing electric fields, but electrons can also be deflected by magnetic fields. The earth's magnetic field is quite weak, so it is unlikely that orienting a valve in any particular direction will affect distortion, but many electrodes are made of nickel, which can easily be magnetized. If the valve was constructed from concentric cylindrical electrodes, magnetic deflection would not matter unless it caused electrons to miss the anode, but box constructions do not have radial symmetry, so horizontal magnetic deflection could influence anode current.

Beam tetrodes with aligned grids are the most susceptible to magnetic fields because vertical magnetic deflection would cause the sheets of electrons to intercept g_2 rather than passing cleanly between the vertically aligned windings. Thus, a magnetic field can change the I_a/I_{g2} ratio, and it would be foolish to suggest that this could not affect distortion. Some years ago, using a coil intended for degaussing television display tubes, the author jokingly degaussed the KT88 (aligned grid beam tetrodes) of a power amplifier, and everyone heard a slight difference.

We should be aware that degaussing requires the magnetic material to be taken to saturation in both directions and then gently taken through ever decreasing hysteresis loops until the residual magnetism is zero. Thus, magnetization, and demagnetization, are achieved by brute force – the author's degaussing coil is 10″ (250 mm) in diameter, consumes 750 VA, and is only rated for intermittent use. Applying an audio signal to an amplifier cannot possibly achieve this effect, no matter how exotic the signal may be.

Testing to find low distortion valves

When designing a power stage, the most important consideration is $P_{a(max.)}$, and the consequent DC requirements force low distortion to be quite low down on the list of priorities. Similarly, low noise input stages demand high *gm*, and because signal levels are so low (otherwise noise would not be an issue), distortion is a very low priority.

To minimize noise, well-designed circuitry amplifies low-level signals once only, and thereafter they are treated as line level signals. Similarly, power stages are used once only to drive the load. By contrast, signals between these

extremes may suffer a great deal of processing, so it is essential to minimize their distortion.

Because of the previous two arguments, low distortion valves are essential for line level processing, but they need not have outstanding *gm* or $P_{a(max.)}$. High μ valves may be undesirable because their design assumed the use of negative feedback to reduce distortion. Sadly, most of the low μ valves were designed for television field scan, so their distortion is distinctly questionable unless individually selected. The remaining valves are medium μ, and have $P_{a(max)} < 5\,W$.

The 6SN7 is widely accepted as a low distortion valve, but how well does it justify its reputation? Bearing in mind that valves were assembled by hand and subject to wide production tolerances, is there a 'best' medium μ valve or manufacturer? This section seeks to answer these questions by reporting on the testing of a selection of medium μ valves under identical conditions.

The test circuit

If we require a low distortion gain stage, this can be achieved by a single-ended stage with an active load, or a differential pair with resistive loads and constant current sink tail. In short, circuit design can reduce distortion, but only to the point where the valve's *irreducible distortion* takes effect. If we are to select valves for minimum distortion, we should focus on their irreducible distortion, since higher levels of distortion can always be reduced by suitable circuit design. Although this complicates matters by requiring us to measure distortion in a stage deliberately designed for low distortion, it has the benefit of enforcing a level playing field.

If later, we use a topology that does not minimize distortion, and valve 'A' sounds better than valve 'B', it is because valve 'A' suits the topology better than valve 'B', *not* because valve 'A' is 'better' than valve 'B'.

As previously mentioned, distortion in a triode amplifier is dominated by the variation of r_a with I_a. Provided that $R_L >> r_a$, the variation of r_a is insignificant, so distortion can be reduced by maximizing R_L. In addition, the valve should pass sufficient anode current to place its operating point well clear of the typical bunching of anode curves experienced at low currents. Accordingly, the valves were tested in a μ-follower circuit passing $\approx 8\,mA$ anode current. In this configuration, the test valve sees $R_L \approx 800\,k\Omega$, which although far from infinite, can be put into perspective by realizing that if a true $800\,k\Omega$ resistor were to be substituted, a $6.4\,kV$ HT supply would be required. Heater filaments were supplied by stabilized DC. See Fig. 3.22.

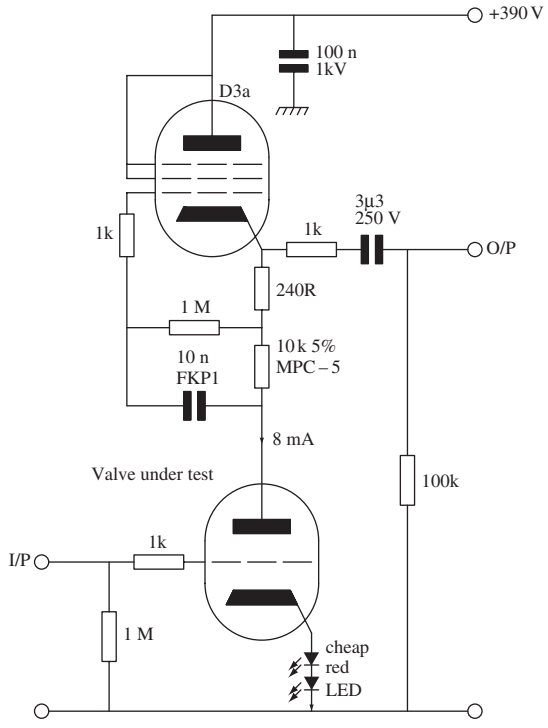


Fig. 3.22 *Medium-μ valve test circuit*

Audio test level and frequency

Since distortion was expected to be low, the valves had to be tested at a sufficiently high output level to make the distortion easily measurable, but well below clipping. +28 dBu($\approx$19.5 V_{RMS}) was found to be a good compromise, so all the valves had their input levels adjusted to produce precisely +28 dBu at their output. Distortion in all the reported valves is directly proportional to level, so distortion at lower levels can be extrapolated from the test data.

Although initially tested at 120 Hz, 1 kHz and 10 kHz, distortion of the test circuit was found to be completely independent of frequency, so the valves were tested at 1 kHz only. For most of the valves, harmonics beyond the 6th were too close to the oscillator distortion residual for reliable measurement, so measurement was only attempted on harmonics up to and including the 6th.

Test results

All of the valves tested were 'New Old Stock' (NOS), so the newest valves are at least 20 years old, and the oldest was 48. Since the valves have been out of production for decades, for some types, only a few samples were available to the author.

Raw data from the measurements were analysed on a spreadsheet, and broken into different groups as and when significant differences became apparent.

The following table summarizes the results of the 6SN7GT/12SN7GT and its direct equivalents. The number of samples refers to the number of individual triodes tested, not envelopes.

Type	**Samples**	**2nd**	σ	**3rd**	σ	**4th**	σ
6SN7GT/12SN7GT	44	−50	3.6	−85	8.4	−96	5.9
7N7	82	−52	3.3	−85	8.6	−97	6.7
14N7	62	−52	3.3	−85	8.6	−97	6.7
Blackened 6SN7GT	6	**−54**	1.8	**−94**	5.6		
Blackened CV1988	12	**−57**	2.6	−85	7.2	−93	4.2
12SX7GT	12	−50	1.9	−83	3.2	−94	6.0
GEC/Marconi B36	6	−51	2.0	**−90**	8.1	−88	2.0
6J5GT (Various)	6	−50	4.1	−82	12.7	−97	3.1
Pinnacle 6J5GT	138	−52	2.6	**−90**	6.7	−96	3.9
RCA 6J5	15	−47	4.8	−84	8.3	−89	7.7
GEC L63	5	−50	1.6	−86	4.4	−89	4.4
7A4	3	−48	0.2	−73	1.6	−93	1.2

(σ = standard deviation)

The next table normalizes distortion to the 6SN7GT/12SN7GT to enable clearer comparison.

Interpretation

The manufacturers claim that all the preceding valves are electrically equivalent. Nevertheless, there are significant differences between the valves, and useful conclusions can be drawn from the second table:

Type	Samples	2nd		3rd		4th	
		dB	ratio	dB	ratio	dB	ratio
6SN7GT/12SN7GT	44	0	1	0	1	0	1
7N7	82	−2	0.79	0	1	−1	0.89
14N7	62	−2	0.79	0	1	−1	0.89
Carbonized 6SN7GT	6	**−4**	**0.63**	**−9**	**0.35**	**−14**	**0.2**
Carbonized CV1988	12	**−7**	**0.45**	0	1	+3	1.4
12SX7GT	12	0	1	+2	1.26	+2	1.26
GEC/Marconi B36	6	0	1	−5	*0.56*	+8	2.5
6J5GT (various)	6	0	1	+3	1.4	−1	0.89
Pinnacle 6J5GT	138	−2	0.79	−5	**0.56**	0	1
RCA 6J5	15	+3	1.41	+1	1.12	+7	2.2
GEC L63	5	0	1	−1	0.89	−3	0.71
7A4	3	+2	1.26	+12	4	+3	1.4

- Valves with carbonized glass envelopes produce less distortion. Deketh reported that carbonized envelopes reduced distortion at maximum power, but this series of tests suggests that the improvement is proportional to level, and that carbonized envelopes significantly reduce distortion at lower amplitudes.
- The RCA 6J5 has significantly higher distortion than the 6J5GT, probably due to increased numbers of gas ions resulting from outgassing of the metal envelope causing increased grid current.
- Despite having a clear envelope, the (Russian made) Pinnacle 6J5GT offers very low distortion – significantly better than any other manufacturer of 6J5GT.
- The Loktal™ base was specifically designed to reduce stray capacitances and inductances by eliminating the glass pinch required by the Octal base, hence the 6SN7GT/12SN7GT has a claimed $C_{ag} \approx 4\,\text{pF}$, whereas the 7N7 has a claimed[6] $C_{ag} = 3\,\text{pF}$.
- Some valves were selected from the standard production line by their manufacturers. This test shows no significant difference in distortion for the 12SX7 (simply a 12SN7GT selected for transconductance[7] at $V_a = 28\,\text{V}$) compared to the standard 6SN7GT/12SN7GT.
- The 3rd harmonic distortion of the Loktal™ 7A4 single triode is very disappointing, but as only three samples were available for testing, the results are not statistically significant.

The *SN7GT family is available with four different heater filament constructions,[8] so the following table compares the different types.

	Voltage	Current	Heating power
6SN7GT	6.3 V	600 mA	3.78 W
8SN7GT	8.4 V	450 mA	3.78 W
12SN7GT	12.6 V	300 mA	3.78 W
25SN7GT	25 V	150 mA	3.75 W

As can be seen from the table, the heater power is almost identical for all types. Physically, the 6SN7GT has its heaters wired internally in parallel, whereas the 12SN7GT has them wired in series, but electrode construction is identical, so distortion ought to be similar. The next table compares distortion of the 6SN7GT with that of the 12SN7GT.

	Samples	2nd	σ	3rd	σ	4th	σ
6SN7GT	28	−50	3.5	−83	8.9	−96	5.7
12SN7GT	16	−51	3.8	−87	7.3	−97	6.5

The two valves are similar – the differences are well within the margins of error.

Similarly, we can compare the 7N7 with the 14N7.

	Voltage	Current	Heating power
7N7	6.3 V	600 mA	3.78 W
14N7	12.6 V	300 mA	3.78 W

Again, we would expect the distortion between the two types to be similar.

	Samples	2nd	σ	3rd	σ	4th	σ
7N7	82	−52	3.3	−85	8.6	−97	6.7
14N7	62	−52	2.4	−88	7.8	−95	6.4

Summarizing, the differences between valves having different heater voltages are well within the margins of error. This is good news because it means that we don't have to use the expensive 6.3 V heater valves, but can use the cheaper and more plentiful 12.6 V heater valves and enjoy the reduction in induced hum caused by their reduced heater current.

A convention

The author has adopted a convention that will be applied throughout this book. Having established that the 6J5GT, 6SN7GT, 12SN7GT, 7N7, 14N7 are electrically virtually identical, and that the 8SN7GT and 25SN7GT can be expected to be similar, economy of nomenclature is necessary. From now on, this family will be called the *SN7/*N7 family (to distinguish between the bases and capacitances). Please note that the 6N7 is a *completely* different beast, and bears no relation to a 7N7.

Alternative medium μ valves

The following table shows possible alternatives to the *SN7/*N7 family.

Type	μ	Samples	2nd	σ	3rd	σ	4th	σ	5th	σ	6th	σ
7AF7	16	4	−38	0.3	−62	1.5	−74	0.6	−89	4.2	−91	5.7
ECC82/12AU7/ B329	18	28	−37		−56	1.4	−73	3.9	−86	6.6	−96	3.1
E182CC/7199	18	30	−45	1.7	−70	1.5	−92	3.7				
E288CC	20	14	−49	1.3	−69	0.9	−89	5.4	−95	7.2	−96	4.9
37	9	9	−45	0.6	−69	4.9	−87	5.7	−88	10.1	−86	14.2
5687 (various)	16	22	−49	1.1	−72	1.7	−91	3.9				
Philips 5687WB	16	14	−42	2.5	−68	2.8	−92	2.4				
6350	20	26	−44	1.4	−65	2.4	−84	2.4	−98	6.2		

For each valve, μ was read from the anode curves at the 8 mA operating point.

The next table allows quick comparison of these alternatives by normalizing their distortion to the 6SN7GT/12SN7GT.

Type	Samples	2nd		3rd		4th	
		dB	ratio	dB	ratio	dB	ratio
6SN7GT/12SN7GT	44	0	1	0	1	0	1
7AF7	4	+12	4	+23	14	+23	14
ECC82/12AU7/B329	28	+13	4.5	+29	28	+23	14
E182CC/7199	30	+5	1.78	+15	5.6	+4	1.58
E288CC	14	+1	1.12	+16	6.3	+7	2.2
37	9	+5	1.78	+16	6.3	+9	2.82
5687 (various)	22	+1	1.12	+13	4.5	+5	1.78
Philips 5687WB	14	+8	2.5	+17	7.1	+4	1.58
6350	26	+6	2	+20	10	+12	4

The results speak for themselves. All of the alternatives are inferior to the *SN7/*N7 family, and produce significantly more 3rd harmonic distortion.

The Loctal 7AF7 dual triode and B9A ECC82 dual triode are particularly ghastly. Perhaps significantly, these valves have an electrode structure that significantly reduced C_{ag} compared to the *SN7/*N7 family (2.3 pF and 1.6 pF vs 4.0 pF). These tests suggest that the measures necessary to reduce C_{ag} within the electrode structure might adversely affect distortion.

There was a significant difference between the Philips 5687WB and other manufacturers' samples, so this type was isolated. Although the 2nd and 3rd harmonics are significantly higher than competing manufacturers, the 2nd would mostly be nulled if used in a differential pair.

Distortion weighted results

Towards the beginning of this chapter, distortion weighting was suggested as a useful technique, so the following table is weighted according to CCIR/ARM. Because this particular test was limited to harmonics up to the 6th (6.3 kHz is the corner frequency for CCIR/ARM), and the distortion was dominated by 2nd harmonic, the difference between CCIR/ARM and the Shorter recommendation was only $\approx$0.1 dB.

Type	Number of samples	Weighted distortion (dB)
Carbonized CV1988	12	−58
Carbonized 6SN7GT	6	−55
Pinnacle 6J5GT	138	−52
7N7/14N7	144	
GEC/Marconi B36	6	−51
6SN7GT/12SN7GT	44	−50
12SX7GT	12	
6J5GT (not Pinnacle)	6	
L63	5	−49
E288CC	14	
5687 (not Philips)	22	
7A4	3	−48
RCA 6J5	15	
E182CC/7199	30	−45
6350	26	−44
Philips 5687WB	14	−42
7AF7	4	−38
ECC82, 12AU7, B329	28	−36

Overall conclusions

A total of 529 valves was tested, and the results show that the reputation of the *SN7/*N7 family is well justified. Distortion for the dual triodes varied from sample to sample, with few significant trends visible between manufacturers. If individual measurement and selection for low distortion are not possible, then carbonized envelope valves from the *SN7/*N7 family are likely to produce the lowest distortion. If these are not available, then a 7N7, 14N7 or a Pinnacle 6J5GT is likely to be a good choice. Valves with B9A bases are significantly poorer.

Coupling from one stage to the next

The most common way of coupling one stage to the next is via a capacitor. A perfect capacitor does not generate distortion. Unfortunately, even a perfect capacitor can exacerbate the distortion generated by valves or transistors.

Blocking

Blocking is an extremely unpleasant mechanism whereby an amplifier mutes for a short time after a momentary overload. In simple terms, blocking is caused by the capacitor that couples to a stage that is overloaded. See Fig. 3.23.

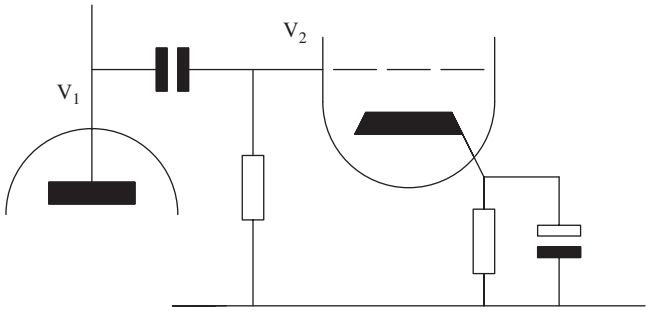


Fig. 3.23 *Capacitor coupling and blocking*

Capacitor coupling between two stages forms a high-pass filter. In order not to affect the audio, we deliberately design for a low $f_{-3\,\mathrm{dB}}$ frequency using:

$$f_{-3\,\mathrm{dB}} = \frac{1}{2\pi CR} \text{ alternatively, } f_{-3\,\mathrm{dB}} = \frac{1}{2\pi\tau}$$

Thus, setting $f_{-3\,\mathrm{dB}} = 1\,\mathrm{Hz}$ means that $\tau = 160\,\mathrm{ms}$, but the true significance of τ in this context will take a little time to emerge.

Because of Kirchhoff's voltage law, the following must be true at *all* times:

$$V_C = V_{a(V_1)} - V_{g(V_2)}$$

In our example, $V_{a(V_1)} = 100\,\mathrm{V}$, and the following grid is tied to ground via a grid leak resistor, so $V_{g(V_2)} = 0\,\mathrm{V}$, causing $V_C = 100\,\mathrm{V}$.

If we apply an impulse to V_1 so that the anode swings 20 V positively, the grid of V_2 attempts to rise to +20 V, but when it reaches +10 V, $V_{gk} = 0$, the grid conducts so heavily to the cathode that its voltage is clamped to +10 V. At this instant, the previous equation must *still* be true, so:

$$V_C = 120\,\mathrm{V} - +10\,\mathrm{V} = 110\,\mathrm{V}$$

The capacitor was able to change its voltage almost instantaneously because it charged through the low impedance path of the overloaded grid.

When the impulse passes, we can find the grid voltage of the second valve by rearranging the equation:

$$V_{g(V_2)} = V_{a(V_1)} - V_C$$

$$V_{g(V_2)} = 100\,V - 110\,V = -10\,V$$

The grid is at $-10\,V$, but the cathode is held at $+10\,V$ by the cathode bypass capacitor, so $V_{gk} = -20\,V$, and the grid has reverted to a high impedance path. More importantly, the valve is now *switched off* and remains so until the grid recovers to $0\,V$. The only path for the capacitor to change its charge is through the grid leak resistor, but as we saw earlier, this combination has a time constant of 160 ms. Worse, 5τ is required for the capacitor to change its charge to within 99% of its final charge, so the grid does not reach $0\,V$ until 0.8 s after the momentary overload.

Recovery from overload is complicated by the fact that once the valve is switched off, there is no cathode current, so the cathode bypass capacitor begins to discharge through the cathode bias resistor. Although this causes the valve to begin conducting earlier than would otherwise have been expected, it also has to recover from the change imposed by blocking. Thus, a momentary overload has had its effects extended to almost a second.

It might be thought that the severe overload posited to cause blocking is unlikely, but applying global feedback around a power amplifier with a capacitor coupled output stage almost guarantees blocking. Suppose that a transient causes clipping of the output stage. Feedback attempts to correct this distortion of the waveform by increasing substantially the drive to the output stage, and the requirements for blocking have been satisfied.

Blocking occurred because a coupling capacitor was allowed to change its charge significantly during an overload. If the capacitor could be eliminated, or moved so that it coupled to a stage that could not be overloaded, the problem could not arise. This technique will be explored in the non-blocking 'Crystal Palace' amplifier described in Chapter 6.

Transformer coupling

Transformers are expensive, but they are essential for connecting loudspeakers to valve amplifiers unless we are prepared to tolerate appallingly low efficiency. Interstage transformers offer some unique advantages.

From the point of view of the primary, if the transformer is used as the anode load, the valve can achieve a much larger signal swing because the anode can theoretically swing to double the HT voltage. Because the transformer effectively doubles the HT voltage, the signal swing is proportionately smaller, which reduces distortion compared with the same valve having a resistive anode load from the same HT voltage. A push–pull stage allows cancellation of even harmonic distortion. Unfortunately, a single-ended stage forces the transformer to pass the (DC) anode current, requiring a gapped core, which reduces bandwidth. If the transformer steps down by a ratio of 2:1, the stage can produce the same swing as a resistively loaded stage, but with a quarter of the output resistance.

From the point of view of the secondary, a centre-tapped winding provides ideal phase splitting. Larger power valves need low grid leak resistances because of their grid current, so the very low R_{DC} of the secondary is ideal.

Against these advantages we have to weigh the inescapable fact that interstage transformers suffer most from transformer imperfections because they operate at high impedances.

Level shifting and DC coupling

In the absence of PNP valves, DC coupled valve amplifiers rely fundamentally on the potential divider. See Fig. 3.24.

In Fig. 3.24a, we have a simple resistive potential divider to the negative HT. We want to achieve $-10\,V$ at the output of the potential divider. Rather than use the potential divider equation, it is easier to set a current through the potential divider and apply Ohm's law to find the required

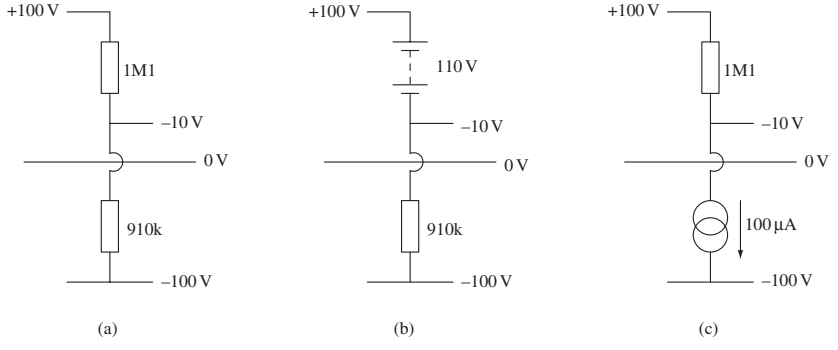


Fig. 3.24 *The three potential divider arrangements that allow DC coupling*

resistances. Our potential divider will steal current from the anode of the preceding valve, so we should minimize this current. If we set the potential divider current to 100 μA, the upper resistor must be:

$$R = \frac{V}{I} = \frac{100\,\mathrm{V} - -10\,\mathrm{V}}{100\,\mu\mathrm{A}} = \frac{110\,\mathrm{V}}{100\,\mu\mathrm{A}} = 1.1\,\mathrm{M}\Omega$$

Similarly, the lower resistor must be:

$$R = \frac{V}{I} = \frac{-10\,\mathrm{V} - -100\,\mathrm{V}}{100\,\mu\mathrm{A}} = \frac{90\,\mathrm{V}}{100\,\mu\mathrm{A}} = 900\,\mathrm{k}\Omega$$

The nearest value of 910 kΩ is fine. Unfortunately, not only have we *level shifted* our signal by the required amount, we have also attenuated it. Thinking in AC terms:

$$\frac{v_{\mathrm{out}}}{v_{\mathrm{in}}} = \frac{910\mathrm{k}}{910\mathrm{k} + 1\,\mathrm{M}1} = 0.453 = -6.9\,\mathrm{dB}$$

Pure resistive level shifters inevitably attenuate, and this is the price that must be paid for simple DC coupling.

Alternatively, we could replace the upper resistor with a battery, to make a Thévenin level shifter. See Fig. 3.24b.

Because the battery is a perfect Thévenin source, it is a short circuit to AC, so this level shifter does not attenuate AC. Because 110 V batteries are inconveniently large, we would replace the battery with a Zener diode or neon reference valve. Sadly, both devices must pass a significant DC standing current (typically 5 mA), which makes them difficult to use. Worse, they are both noisy.

Our final possibility is to replace the lower resistor with a constant current sink, to make a Norton level shifter. See Fig. 3.24c.

There is no reason why we should not make a constant current sink out of bipolar transistors or a pentode. Provided that the sink has $r_{\mathrm{out}} >> R_{\mathrm{upper}}$, the Norton level shifter does not attenuate AC. However, there is a noise problem. Pentodes and transistors are *transconductance amplifiers*, which means that they convert their input voltage into an output current. A current sink amplifies its DC reference voltage, and we are now proposing to convert

its output current back to a voltage using a high value resistor. In effect, we have constructed a high gain amplifier that amplifies the noise voltage of the DC reference.

Although Norton level shifters are usable, their major problem is noise, and almost all the design effort must focus on reducing their noise to acceptable levels. (Constant current sink tails in differential pairs do not add significant noise because they see the low impedance load r_k, and their noise is common mode, so it is mostly rejected.)

Sadly, all of these techniques connect the signal to the negative HT, which is also capable of adding hum and noise to the wanted signal.

A DC coupled Class A electromagnetic headphone amplifier

As with all amplifiers facing a difficult load, this circuit was designed by working backwards from the output to the input. Be aware that the circuit is presented as a vehicle for solving the problems of DC coupling – *not* as a detailed exemplar of headphone amplifier design.

A cathode follower is needed to provide a low output resistance. High μ, high *gm* valves are best suited as cathode followers because the high *gm* ensures low r_{out}, and high μ allows plenty of feedback to reduce distortion, so the 6C45Π is ideal. The cathode is connected to the negative HT via a standard pentode constant current sink using an EL822.

We know that we want to apply DC feedback to stabilize the output of the amplifier at 0 V, and we need our completed amplifier to have a high input resistance so that it does not load volume controls. The ideal input stage is therefore a differential pair, and because we already have a negative HT, it seems churlish not to use another pentode constant current sink. See Fig. 3.25.

Electromagnetic headphones are low impedance devices. Because portable equipment must operate from a 3 V battery (possibly only 1.5 V), headphones designed for portable use are typically 32 Ω, but better quality designs tend towards 200 Ω. Either way, they require considerable current, and are extremely onerous loads for valves. As we increase I_a, P_a rises, so we must reduce P_a by lowering V_a as far as we dare without running into grid current. Setting $V_a = 135$ V keeps us just clear of grid current. Although it is claimed that $P_{a(max.)} = 7.8$ W, all of the other 6C45Π specifications are somewhat optimistic, and the envelope is little larger than an ECC88, so it does not seem wise to operate at the full claimed power. If we set $I_a = 34$ mA, then $P_a = 4.6$ W.

We now need to consider the input differential pair. Because we only have 135 V of HT, we need a valve having good linearity at low voltages. The ECC86 would be ideal, but the ECC88 was available, and a little slithering

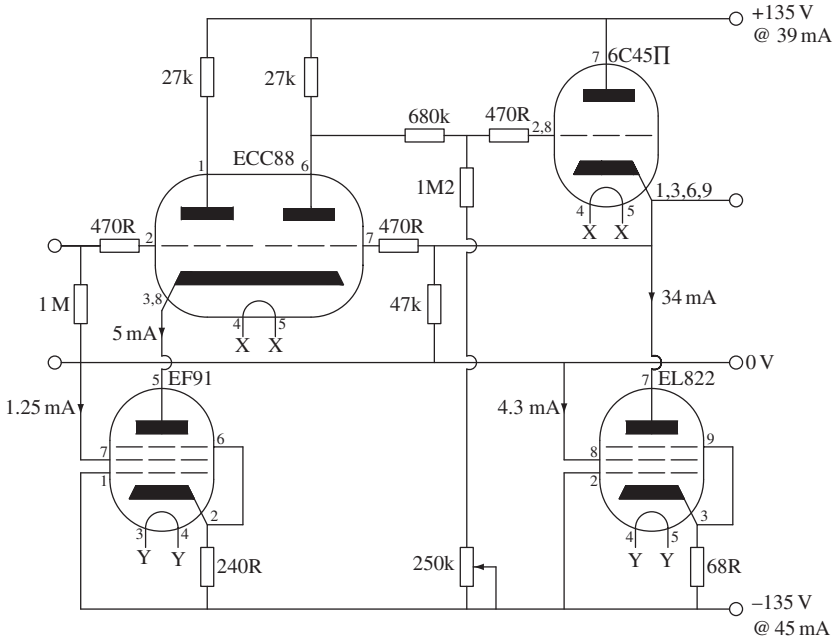


Fig. 3.25 *A DC couped amplifier using a potential divider to the negative HT*

of loadlines suggested that 27 kΩ load resistors would work well with an anode voltage of 68 V, resulting in a tail current of 5 mA at $V_{gk} = 2$ V.

We are finally in a position to be able to investigate the more significant problems of DC coupling using a pure potential divider.

Each ECC88 anode is at 68 V. The grid of the 6C45Π cathode follower is at ≈ -1.5 V (remember that we need the cathode to be at 0 V), so we must drop 69.5 V. If we were to set 100 μA of potential divider chain current, we would need an upper resistor of 695 kΩ, which is an awkward value. If we choose the nearest preferred value of 680 kΩ, our potential divider current becomes 69.5 V/680 kΩ = 102.2 μA.

The lower resistor must drop $-1.5\,\text{V} - -135\,\text{V} = 133.5\,\text{V}$, so 102.2 μA requires a 1.3 MΩ resistor, which *is* a preferred value. However, we need to consider another factor.

At the calculated anode voltage, the potential divider chain steals ≈100 μA current from the anode current of 2.5 mA. 2.5 mA is a low current for this valve, so sample variations between valves are greatly magnified, and the small stolen current is negligible by comparison. The theft/valve error pulls the output DC in the final circuit away from 0 V, so why not make

the divider chain adjustable to compensate for valve variation? Using a 1.2 MΩ fixed resistor in series with a 250 kΩ resistor allows ±10% variation. You could apportion a larger value of variable resistance, but this would make the output DC adjustment more fiddly.

Even if the output DC is carefully adjusted to 0 V, it will drift. We need a means of forcing the output DC to 0 V, and the best way of doing this is to apply negative feedback. We connect our feedback network in *parallel* with the output of the amplifier, but because the output of the feedback potential divider is connected to the other grid of the differential pair, and the differential pair amplifies the *difference* between its inputs, it is in *series* with the input signal. The feedback is therefore parallel derived and series applied, so it reduces output resistance and increases input resistance. Reduced output resistance is significant because all contemporary electromagnetic transducers rely on $r_{source} = 0$ to produce their optimum transient response.

If we are to apply feedback, we must know how much gain is available. 27 kΩ is quite a low anode load resistance for an ECC88, and a loadline predicts a gain of 26.75. The input stage operates as a differential pair, but because only one output is used, we must immediately halve the gain to 13.375. We incur a loss of 0.657 in the level shifter, which reduces the gain to 8.78. Considering the 6C45Π, at $I_a = 34$ mA, $r_k \approx 25\ \Omega$, so a 32 Ω load incurs a further loss of 0.56, leaving a total gain of ≈5. Thus, even 100% global feedback could produce only $(1 + \beta A_0) = 5$, or 14 dB, of improvement.

The circuit was tested with 100% negative feedback because this is the most critical condition for stability, yet delivered an exemplary square wave response at 10 kHz into a 200 Ω load. See Fig. 3.26.

The circuit was tested into various load resistances.

	Output in dBu for specified THD + N		
	0.5%	**0.2%**	**0.1%**
100 kΩ	–	–	+20
200 Ω	+16.7	+12.8	+7.8
32 Ω	−3.7	–	–

Unsurprisingly, the table shows that little valves don't like 32 Ω loads, and that this circuit can only deliver 8 mW at 0.5% THD + N into a 32 Ω load.

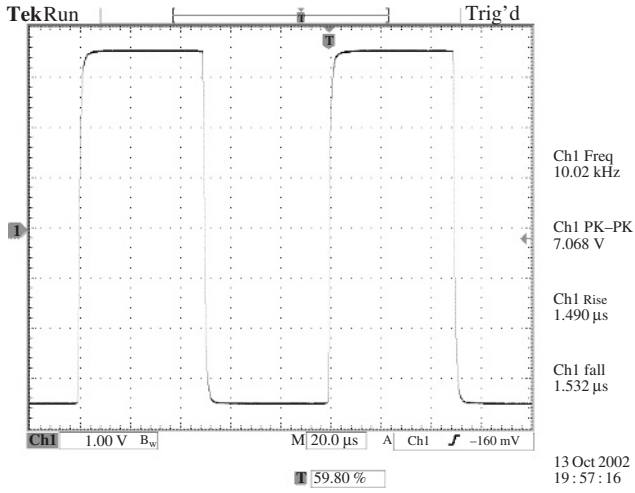


Fig. 3.26 *Near-perfect 10 kHz square wave response*

The performance improves markedly into a 200 Ω load, whereupon the circuit delivers double the power at <0.1% THD + N. More importantly, the spectrum of the distortion harmonics becomes acceptable: 2nd = −60 dB, 3rd = −82 dB, 4th = −100 dB. Curiously, the author does not own any electromagnetic headphones of plausible quality, so he was unable to test this amplifier subjectively.

Eagle-eyed readers having a 6C45Π data sheet will notice that this circuit far exceeds the maximum specified grid leak resistance of 150 kΩ. Remember that the purpose of the grid leak resistor is to hold the grid at its correct voltage in the face of grid current. If the resistor is too large, grid current raises the voltage on the grid, reducing V_{gk}, increasing I_a, causing more grid current, until the valve dies. In this circuit, I_a for the 6C45Π is set purely by the constant current sink, so considerations of thermal runaway caused by excessive grid leak resistance are irrelevant.

If you need to set up a circuit of this type, it is easiest to break it into two parts before applying feedback. Build the output stage first, short circuit the upper grid to ground, and adjust the CCS current programming resistor to set correct cathode follower current. Build the differential pair and associated CCS, and adjust the CCS current programming resistor to set required anode conditions. Connect the two stages, and fine-tune the level shifter for 0 V at the output. Finally, close the negative feedback loop.

Using a Norton level shifter

As mentioned previously, the Norton level shifter is a high gain amplifier that amplifies the noise of its reference voltage. The noise problem can be tackled in various ways:

- Reduce the noise produced by the reference. Forward biased diodes produce less noise, so cheap red LEDs are ideal. If a Zener diode must be used, it should be filtered.
- Noise isn't a problem in itself, it becomes a problem when the signal voltage is sufficiently low that the signal to noise ratio is compromised. Solution: don't use Norton level shifters in pre-amplifiers.
- If the noise can be arranged to become common mode, then it can be rejected by a differential pair. This is the most powerful technique.

Some years ago, the author picked up 40 N-channel and P-channel MOSFETs in a junk shop, and after passing them through the curve tracer recently, he found two reasonably complementary pairs. The hybrid idea promptly resurfaced, coupled with a wish to use some of the more obscure valves in stock. The author had previously been unable to find a use for the wonderfully high *gm* (55 mA/V) of the E55L, but realized that it might make a good cathode follower for driving the high gate capacitance of MOSFETs. A little doodling resulted in a circuit requiring a Norton level shifter. See Fig. 3.27.

The N-channel MOSFET needs +5 V on its gate to pass the required current of 1.7 A, whilst the P-channel MOSFET requires −6 V, so the V_{be} multiplier between the gates allows these voltages to be applied and output stage current to be set. The E55L cathode follower has a power cascode constant current sink made from an MJE340 and a BC549 as its active load. The 7N7 differential pair has a cascode constant current sink tail that shares the reference voltage of the sink for the E55L. In order to balance the anode loads, the unused output of the 7N7 has an RC network to ground to simulate the input impedance of the E55L cathode follower. The ECC808 input differential is reasonably conventional except that it has constant current source anode loads to improve its linearity when operating from a positive HT of only +150 V (its cathode CCS may need fine adjustment to set V_a).

Because the 7N7 differential pair is directly coupled to the E55L cathode follower, its grids must be at, or close to, the voltage of the negative HT, yet the anodes of the ECC808 are expected to be at +123 V. Thus, the problem is to DC couple the two stages with a minimum of noise.

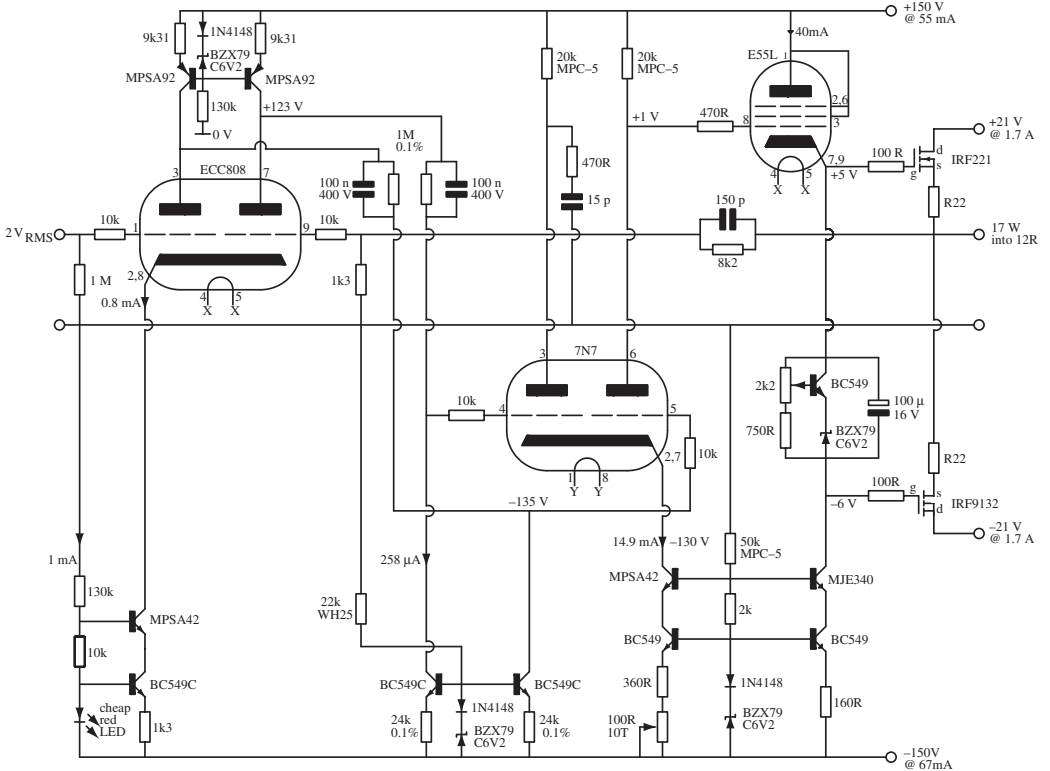


Fig. 3.27 *DC coupling using a Norton sink*

Arbitrarily selecting 1 MΩ for the upper resistor of the level shifter suggests that the current passing through it will be ≈250 μA. We need to know this current because it is the design current for the constant current sink. If we choose a 6.2 V reference voltage, 24 kΩ is required. If we treat the level shifter as a common emitter amplifier, we can find its gain. $I_c = 250\,\mu A$, so $gm = 35I_c = 35 \times 0.25 = 8.75\,mA/V$. The gain of the amplifier is $A_v = gm \cdot R_L = 8.75 \times 1000 = 8750$. However, the amplifier has considerable feedback because of the unbypassed 24 kΩ emitter resistor, so we can use the feedback equation:

$$A_v = \frac{A_0}{1 + \frac{R_e}{R_c} \cdot A_0} = \frac{8750}{1 + \frac{24}{1000} \cdot 8750} = 41.5$$

Alternatively, now that we know that the gain before feedback is likely to be large, we can simply use the approximation:

$$A_v = \frac{R_c}{R_e} = \frac{1000}{24} = 41.7$$

The significance of this exercise is that a larger reference voltage reduces the gain of the amplifier because of the high value of R_e that it enforces. Although an LED would be quieter, it would cause the amplifier to have a gain of ≈250.

If the level shifters share the (noisy) DC reference, the noise is amplified identically, so it is presented to the differential pair as common-mode noise, which it can reject.

To ensure that the noise remains common mode, the emitter resistors and 1 MΩ need to be matched, so 0.1% tolerance types are ideal. At high frequencies, the common-mode rejection of the differential pair deteriorates. If we bypass the 1 MΩ resistors with capacitors, the gain of the level shifter will fall with frequency, so it presents less noise to the differential pair, which compensates for the falling CMRR of the differential pair.

References

1 The Influence of High-Order Products in Non-Linear Distortion. *Electronic Engineering*, April 1950, pp. 152–153.
2 Henderson F. E. (GEC) (1943) Introduction to Valves. 2nd ed. *Wireless World*.
3 Allen Wright (1994) *The Tube Pre-amp Cookbook*.

4 US Patent No. 2 102 671 Harold S. Black.
5 Deketh J. (1949) *Fundamentals of Radio-Valve Technique*. Philips Gloeilampenfabrieken (Reprinted by Audio Amateur Press 1999), pp. 439–444.
6 *Technical Manual Sylvania Radio Tubes*, 6th ed. (1946) Sylvania Electric Products Inc. Emporium, Pennsylvania.
7 Electron Tubes, Receiving Types 12SN7GT and 12SX7GT, MIL-E-1/63B, 4 March 1954 (USA military specification sheet).
8 Piotr Mikolajczyk and Bohdan Paszkowski (1994) *Electronic Universal Vade-Mecum*, 2nd ed. Wydawnictwa Naukowo-Techniczne, p. 442.

Further reading

Metzler, Bob (1993) *Audio Measurement Handbook*. Audio Precision.
Self, Douglas (2000) *Audio Power Amplifier Design Handbook*, 2nd ed. Newnes.
Talbot-Smith, Michael (2002) *Audio Engineer's Reference Book*. Focal.

4

Component technology

In Chapter 2, we began to design simple circuits, which will later be combined to form complete systems. In doing this, we calculated values for components. We now need to know how to specify voltage or thermal ratings of components.

Correct specification of individual components is extremely important. An underspecified component may fail prematurely, and cause further damage as it does so, whereas an overspecified component may waste money that could have made an improvement elsewhere. To be able to specify components correctly requires knowledge of the stresses that will be applied to the component (electrical, thermal and mechanical), and knowledge of the imperfections of that breed of component. (No components are perfect, although some are more equal than others.)

Much has been said about the 'sound' of components, particularly capacitors. This has caused such polarization of the Engineers vs Audiophiles debate that rational speech has only rarely been heard. This is curious, since there are well-known physical imperfections in components, and it seems reasonable to suppose that they could have an influence on sound quality. On the other hand, whilst components are not magical, there *are* purveyors of snake oil.

This chapter will help you to avoid the more obvious pitfalls, but it is *not* a substitute for detailed manufacturers' data sheets and the application of intelligence.

Resistors

Preferred values

So far, we have calculated resistor values, and then picked the nearest preferred value. These preferred values are known by their *E series* (E6,

E12, E24, and E96), whose values are given in the Appendix. Each series denotes the number of different values in a decade.

For example, E6 contains the values 1, 1.5, 2.2, 3.3, 4.7 and 6.8, making a total of six values per decade. If we now consider that we will probably need values from 1 Ω to 10 MΩ, then this is seven decades, and we will need 43 different values (10 MΩ is the start of a new decade). For a complete set of E24 (the most commonly used range) we would need 169 different values.

The E series is loosely related to the tolerance of the component, so 20% tolerance components are E6. The reason is that the upper limit of tolerance of one value just meets the lower limit of tolerance of the next highest value, so there are no gaps in the range.

The argument begins to fall down when we look at E24, since $1.3 + 5\% \neq 1.5 - 5\%$.

Heat

Resistors convert electrical energy into thermal energy. The amount of energy converted per second is the power, and this determines the temperature rise. A signal resistor is unlikely to be a problem, but an anode load resistor could dissipate significant power. We can easily calculate the power dissipated as V^2/R, and select an appropriate component.

This is not actually quite as easy as it sounds, and there is plenty of scope for getting it wrong.

Resistor manufacturers typically specify power ratings at a component temperature 70°C (158°F).

If your equipment operates in a typical domestic temperature of 20°C (68°F), then the *internal* temperature must be higher than this, since the equipment is consuming energy and is not 100% efficient. An average internal temperature of 40°C (104°F) would be quite likely, whilst areas of localized heating (hotspots) could be considerably higher. If you are fortunate enough to live somewhere warmer, then an external temperature of 35°C (95°F) might not be unusual, and the internal temperature would rise accordingly.

We can only lose heat from a higher temperature to a lower one, and we can make a useful electrical analogy.

Temperature difference, ΔT (°C) is equivalent to potential difference.

Power dissipation q (W) is equivalent to current.

Thermal resistance R_θ (°C/W) is equivalent to electrical resistance.

From this we can derive a thermal 'Ohm's law':

$$\Delta T = R_\theta \cdot q$$

This tells us that a given thermal resistance will create a greater temperature rise above ambient as more power is dissipated. Resistor specifications give a value for the thermal resistance R_θ, but this value assumes that the flow of air to cool the resistor by convection is *not restricted*.

In practice, we often mount the resistor on a printed circuit board (PCB), which considerably restricts the flow if the board is mounted horizontally. Even mounted vertically, there may still be large components, such as capacitors, that block the air flow.

Combining the arguments of restricted air flow and high ambient temperature, it is not generally advisable to operate resistors at more than one third of their 70°C rating unless you are able to do a detailed thermal analysis. Even with this proviso, a resistor operated at one third of its rating will be significantly warmer than its surroundings, so if it changes its temperature, we should expect its electrical parameters to change too. And they do.

Electrical resistance varies with temperature in accordance with the temperature coefficient of the resistor, generally given in parts per million change of value per °C. This may sound small, but a 30°C rise in temperature can cause a significant change in value. Therefore, if we have gone to the expense of using 0.1% resistors in a critical part of a circuit, we should not allow *any* significant power to be dissipated in them if we want their value to remain substantially the same. Maximum dissipation of one eighth of full rated power would not be unreasonable. In addition, we should ensure that the resistor is not heated by other components.

Resistors are available in two main types: metal film resistors and wire-wound. Despite their recent minor cult status, carbon film resistors are an anachronism and will not be considered, as their tolerance and noise specifications are so very poor, although their reduced inductance makes them useful as grid-stoppers.

Metal film resistors

The control of the quality of the materials and processes in the manufacture of metal film resistors determines their performance, so it is worth detailing their construction.

The process starts with the individual insulating ceramic rods onto which the resistive film is to be deposited. These rods must have a smooth surface, as excessive surface roughness varies the thickness of the resistive layer and

causes discontinuities that produce electrical noise. Although the ceramic material is chemically inert, it may have picked up surface contamination, such as grease or packaging materials, so this is burnt off by passing the rods through an oven at a temperature of $>1000\,^{\circ}C$.

Whilst the rods are still hot, they are transferred to a drum in batches of up to 50 000 at a time. The drum is in a high vacuum sputtering system, which is effectively a large valve. An electron gun fires sufficiently high velocity electrons at the nickel–chromium anode (known as the target) that surface molecules are dislodged to form a nickel–chromium vapour. Tumbling the rods in the drum causes the vapour to deposit evenly upon them. The duration in the drum determines the thickness of the film, and this is the first process that determines the resistance of the final resistor.

The thickness of the applied film affects the noise of the final resistor, with thinner films being noisier than thicker films. If the nickel–chromium alloy contains impurities, it will become more granular, and this also causes noise. If the adhesion of the film to the ceramic rod is poor, the film will lift, causing noise, instability and open circuit failure.

A simple nickel–chromium film cannot achieve a temperature coefficient of 5 ppm unaided, but proprietary techniques can improve this by adjusting the chemistry of the film if necessary.

End caps are fitted next, to allow connection to the resistive film. These end caps are an interference fit onto the rods, and their precise fitting is critical. If they are too tight, then as they are pushed on, they will damage the film, but if they are too loose, then they will not make a good contact. Either of these defects causes noise in the finished resistor. Because the end cap is of a dissimilar metal to the resistive film, the interface between the two is a thermocouple, which can generate noise (as thermal EMF), so the cap material must be carefully selected.

Commonly, ferrous end caps are used, but some manufacturers, such as MEC Holsworthy, use non-magnetic end caps on their Holco range, and it is thought that this may be a contributory factor to their good sound. If necessary, end caps or leads of any component can be checked with a small magnet. Sadly, many components use steel-cored leads because the poor thermal conductivity (compared to copper) reduces the temperature rise in associated wiring.

Sputtering is not a particularly precise process, and the resistance of the films typically has a spread of $\pm 10-20\%$. Now that the end caps have been fitted, it is possible to measure this resistance and grade the rods into batches. The purpose of this is to ensure consistency of helixing (see later) and hence so the performance of the product.

Although the rods now have a resistive element, it is quite low resistance, and this must be increased. This is done by cutting a helix through the film from one end cap to the other in order to lengthen the resistive path, whilst at the same time making it narrower. If there are more turns to the helix, this makes a longer, narrower path, and the resistance of the final resistor is proportionately higher, so resistor manufacturers call this parameter *gain*, and we will return to this later.

Traditionally, the helix was cut by a diamond edged circular saw whose depth of cut was critical. If the cut was too shallow, then the resistive film would be incompletely removed, leaving traces of material bridging adjacent turns of the helix. If the cut was too deep, the saw would be damaged on the ceramic rod, and subsequent resistors would be cut poorly. Either defect caused noise in the finished resistor.

The modern technique is to use a YAg laser to cut the helix, which produces a narrower, more precise cut, but even this process is not without its pitfalls. If the energy directed by the laser is insufficient, the resistive layer is incompletely burnt away, causing bridging. If the energy received from the laser is too great, then the resistive film at the edges of the cut becomes disrupted and has an uneven edge. Both defects cause noise.

As the gain of the resistor rises, the track narrows, causing edge imperfections to become proportionately more significant. This is reflected by manufacturers' published noise performance, which shows that the excess noise generated by film resistors rises for values $>100\,k\Omega$. This effect is particularly noticeable for resistors of low power rating, because their smaller physical size demands a higher gain for a given value.

Film resistors also have a maximum *voltage* rating which is independent of their power rating, but is determined by the maximum allowable potential across the gap between adjacent turns of their helix. As the applied voltage rises, it becomes more likely that *tracking* (intermittent voltage dependent conduction) will occur across the gap due to imperfect removal of the film in the gap. Taken to its extreme, a sufficiently high applied DC causes arcing between turns of the helix and permanently damages a film resistor. When using film resistors as anode loads, it is not sufficient simply to ensure that the power dissipation is satisfactory, the voltage rating must also be checked. Typically, higher power components have higher voltage ratings and lower excess noise.

At much lower voltages, tracking is partly responsible for the inclusion of an excess noise specification for the resistor, which is typically given in terms of μV of noise per volt of applied DC. For minimum noise with film resistors, the applied DC across them should be minimized. A typical value for this

excess noise is 0.1 μV/V, which is −140 dB; however, this can rise to 1 μV/V or more for higher resistance values, which would then be −100 dB.

By this means, applying a *signal* voltage across a film resistor generates a signal level dependent noise or *modulation noise*. Since amplifiers contain many resistors, modulation noise could conceivably rise above the thermal noise floor in a low noise amplifier, but be masked by a poorer amplifier.

Laser cutting of the film resistor produces a precisely toleranced resistor, which then has tinned copper wire leads welded to its end caps before being coated with an insulating protective epoxy film onto which the value is marked.

It will be seen that almost every process can cause noise if carried out incorrectly, so resistor manufacturers routinely measure noise or 3rd harmonic distortion as a means of quality assurance. Unfortunately for audio designers, their noise measurement generally uses a 1 kHz bandwidth filter centred on 1 kHz rather than a 20 Hz–20 kHz audio band filter. Nevertheless, the figure is a useful guide to product ranges from a given manufacturer.

The resistor need not have end caps and leads fitted. Surface mount resistors have their ends plated with a silver palladium alloy. When soldering surface mount resistors, it is essential to use a silver loaded solder to prevent the silver leaching out from the plating and reducing solderability.

Metal film resistors are commonly available in values from 1Ω to 10 MΩ, although values up to 50 GΩ are now available off the shelf.

Wirewound resistors

Wirewound resistors are generally designed to dissipate significant power, with 50 W components being readily available, but ratings up to 1 kW are possible. Resistance values cover as many decades as metal film resistors, but the maximum value available is typically 100 kΩ.

Again, the process begins with a ceramic rod as a former for the resistive element, but this time, resistance wire or tape is wound onto the rod and welded to the end caps, to which leads are then welded. Smaller components (< 20 W) are then coated with a ceramic glaze to prevent movement of the windings and also to seal the component. Larger components may have screw terminal end caps and be fitted into an aluminium extrusion to conduct the heat from the resistive element to an external heatsink. However, high value resistors require many closely spaced turns of fine resistance wire, so the possibility of tracking between adjacent turns defines a voltage rating which can easily override the power rating.

Aging wirewound resistors

Scroggie[1] pointed out that because the resistance wire is wound under tension to ensure a consistent wind, this sets up strains within the wire that relieve with time, causing the resistor's value to change. He further suggested that the process could be accelerated by heating the resistors in an oven heated to 135°C for 24 hours. The author tested the theory by measuring his entire stock of aluminium clad wirewound resistors, leaving them in the kitchen oven on its minimum setting for a day, allowing them to cool slowly in the oven, and measuring them again. Even a 3½ digit DVM was able to show significant differences; resistors more than four years old showed no change, but the newest resistors changed by up to ½% in value. It therefore seems sensible to age wirewound resistors intended for anode loads in differential pairs *before* matching.

Noise and inductance of wirewound resistors

Because the resistive element in film resistors is a thin track, they develop excess noise proportional to the DC voltage drop across them ($\approx 0.1\,\mu V/V$). By contrast, surface imperfections of the resistance wire in a wirewound resistor form a very small proportion of its cross-sectional area, and excess noise is virtually non-existent, making them ideal as anode loads in low noise pre-amplifiers.

Wirewound resistors are wound as a coil, and even though $\mu_r \approx 1$ for the ceramic core (making it comparable with an air core), all coils have inductive reactance which might conceivably be significant compared with the resistance.

The resistance of a conductor is:

$$R = \frac{\rho L}{A}$$

where: ρ = resistivity of conductor
L = length of conductor
A = cross-sectional area of conductor

But the wire is of circular cross-section, and the area of a circle is:

$$A = \frac{\pi d^2}{4}$$

Substituting:

$$R = \frac{4\rho L}{\pi d^2}$$

To make resistors cheaply, the resistance wire is wound onto standard sized cores. To ensure efficient heat transfer to the surroundings, and to reduce the possibility of hotspots, the core is completely covered with one layer of wire from end to end with an infinitesimal gap between turns. The number of turns of wire to completely fill a core of length C is:

$$n = \frac{C}{d}$$

The length of this wire is:

$$L = \pi n D = \frac{\pi C D}{d}$$

Substituting into the resistance equation, π cancels, and the resistance achievable by a single layer wirewound resistor is:

$$R = \frac{4\rho C D}{d^3}$$

Simplifying:

$$R \propto \frac{1}{d^3}$$

Inductance is proportional to n^2, and since n is proportional to $1/d$:

$$L \propto \frac{1}{d^2}$$

As observed earlier, it is the *ratio* of L to R that is important, not the absolute value, so:

$$\frac{L}{R} \propto \frac{\frac{1}{d^2}}{\frac{1}{d^3}} \propto d$$

This result is very significant because it shows us that L/R rises as we use thicker wire, so we can expect low value wirewound resistors to possess significant inductance. This theory was tested on a component analyser, which produced models for a selection of wirewound resistors. Because the resistors were aluminium clad, transformer action to the shorted turn might be expected to reduce inductance, but later dissection showed that the coil diameter was half the internal cladding diameter, implying loose coupling and insignificant transformer action. See Fig. 4.1.

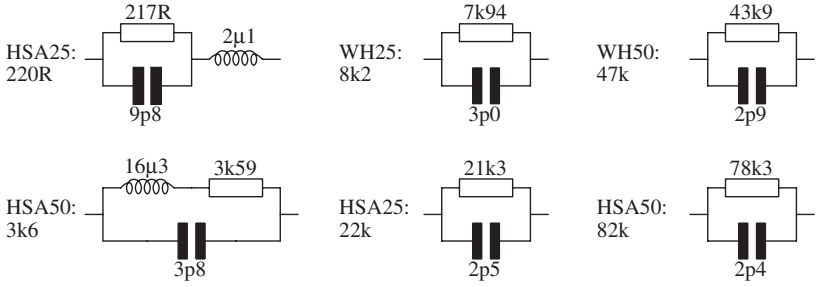


Fig. 4.1 *Equivalent circuits of practical wirewound resistors*

As can be seen from the models, measurement confirms the theory by showing that only low resistance wirewound resistors have significant inductance. Besides deriving models, each resistor was swept from 100 Hz to 100 kHz whilst measuring phase deviation from a perfect resistor. Only the 220 Ω resistor showed a measurable deviation (0.2°).

All of the resistor models required a small shunt capacitor, and once the resistor values were typical of anode load resistances, this shunt capacitor settled to a value of 3 pF ± 1 pF, a value commensurate with the strays that one would expect to find in a practical circuit.

Summarizing, the inductance of wirewound resistors is entirely negligible once their value is greater than 10 kΩ, but, as predicted, inductance becomes significant as the resistance falls. This outcome is very fortunate, since in a typical valve stage, $R_L > 10\,k\Omega$, and requires a power resistor, whereas the cathode bias resistor, $R_k < 1\,k\Omega$, but dissipates little power, and can therefore be a non-inductive component normally intended for component bridges.

General considerations on choosing resistors

Tolerance

- Is the *absolute* value important? If the resistor is part of a network that determines a filter or equalization network, then we need a close tolerance (perhaps even 0.1%) component to minimize frequency response errors.
- Matching: is the component part of a pair? Anode loads in differential pairs should be matched, and so should corresponding components in filter networks for each stereo channel.

Heat

Will the resistor be heated by other components? Will its value change? Will this matter?

Voltage rating

- Is the voltage rating of the component adequate, even under conditions of maximum signal? (Grid leak resistors for low μ power valves, such as 845, might need to consider this factor.)
- Will the DC voltage drop across the resistor develop an unacceptable level of excess noise? If so, a wirewound or bulk foil type should be considered.

Power rating

Is the power rating of the component adequate under all conditions? Could the (varying) *audio* signal heat the resistor sufficiently to change its value and cause an error? If a power component is required, what provision has been made to lose the heat that this component will generate? Will it heat other, sensitive, components?

Capacitors

Capacitors store charge. This charge is stored in the electric field between two plates having a potential difference between them. If there is no potential

difference between the plates, then there is no stored charge, and the capacitor is said to be discharged.

Capacitors for electronic circuits are made of two fundamental components, a pair of conducting plates, and an insulating material called the *dielectric* that separates them. In its simplest form, a capacitor could be a pair of parallel plates separated by a vacuum.

The parallel plate capacitor

Unsurprisingly, the capacitance of a parallel plate capacitor is proportional to the area (A) of the plates, and inversely proportional to the distance (d) between the plates. We should expect this, since if we move the plates an infinite distance apart, they can no longer 'see' one another, and a plate on its own is not much of a capacitor. If the charge is stored between the plates, then it is reasonable to suppose that the interposition of any material (k) between the plates will affect the capacitance. We can formalize these arguments by combining them into a proportionality:

$$C \propto \frac{Ak}{d}$$

To calculate real values in electronic units, we must add some fudge factors, to generate the equation below:

$$C = \frac{A \cdot \varepsilon_0 \cdot \varepsilon_r}{d}$$

This equation looks a lot more impressive, but ε_0 is simply a fudge factor to make the real world fit into our system of units, and is known as 'the permittivity of free space'; it has a measured value of $\approx 8.854 \times 10^{-12}$ F/m. ε_r (also known as 'k') is the *relative* permittivity of the material we insert as the dielectric, compared to the value for a vacuum, and is always >1.

A quick calculation using this equation shows that a parallel plate capacitor in a vacuum (although air is almost identical) with plates 1 m^2, separated by 10 cm, would have a capacitance of 88.5 pF. If we are going to make practical valve amplifiers, we are going to have to do something about the size of this capacitor.

Reducing the gap between the plates and adding plates

One obvious method of increasing capacitance is to reduce the gap between the plates, so typical commercial capacitors use gaps of 5 μm or less.

Another method is to add more conductive plates, in the form of a stack with alternate plates connected together. This almost doubles capacitance over what might at first be expected because we now use *both* sides of each plate (except for the two outermost plates). This form of construction is used for silvered mica capacitors and for stacked film/foil capacitors. See Fig. 4.2.

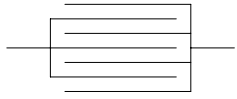


Fig. 4.2 *Cross-section of general form of parallel plate capacitor*

Cutting squares of dielectric and plates and assembling them to form a capacitor is an expensive business, so most capacitors are constructed by winding two long strips of plates and dielectric together to form a cylinder, and then connecting a wire to each plate.

The dielectric

Maintaining a precise air gap of 5 μm between a set of plates would be virtually impossible, so an insulating spacer is needed. This insulating dielectric will have $\varepsilon_r > 1$, which further reduces the physical size of the capacitor for a given value of capacitance.

Unfortunately, we gain this increase of capacitance at the expense of other parameters, and so we should investigate these. The dielectric has three important properties: relative permittivity ε_r, *dielectric strength* and *dielectric loss*.

Relative permittivity, ε_r, has been mentioned earlier, and is effectively the factor by which the capacitance of a capacitor is increased by the insertion of the new dielectric.

Dielectric strength refers to the maximum field strength, measured in volts per metre, that can be applied to a given insulator before it breaks down and conducts. It is this limit that sets voltage ratings for capacitors.

Dielectric loss refers to how closely the dielectric approaches a perfect insulator at voltages *below* breakdown. One way of specifying this loss is to measure the leakage current, in μA, that flows when the maximum rated voltage is applied across the capacitor – this method is typically used for

aluminium electrolytic and tantalum capacitors. Film capacitors are typically rather less lossy, and so the *insulation resistance* or *leakage resistance* of the capacitor may be specified. Dielectric loss may be different for AC to DC, and so a more useful measurement is to measure *tan* δ, which is the ratio of the total resistive component of the capacitor to the reactive component at a specified frequency or frequencies. Note that tan δ does not distinguish between the parallel leakage resistance of the dielectric or any series resistance, such as lead or plate resistance.

Lead and plate resistance are collected together as one term, and are known as the *effective series resistance* (ESR). In components such as high capacitance electrolytic capacitors for power supplies, or cathode bypasses, the ESR is highly significant, since it may be an appreciable fraction of the total impedance of the capacitor. In power supplies, significant currents flow in the reservoir capacitors, which cause self-heating of the internal structure. For this reason, a parameter is quoted that is very closely linked to ESR, and this is *maximum ripple current* (see Chapter 5).

The leads have series inductance, and unless precautions are taken, the plates also have inductance. We can now draw a simple equivalent circuit for a real capacitor. See Fig. 4.3.

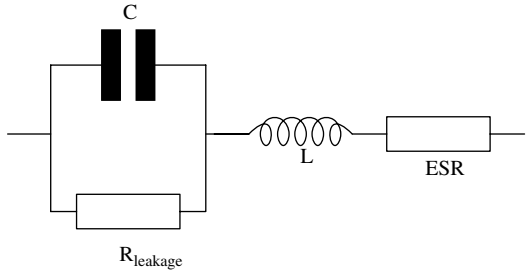


Fig. 4.3 *Basic equivalent circuit of practical capacitor*

It is immediately apparent that we are dealing with a resonant circuit, and for electrolytic capacitors, this self-resonant frequency is often specified in the manufacturers' data sheets, and we will return to this later.

Different types of capacitors

With the various ways of making the plates or the dielectric, there are many combinations of capacitor construction available. See Fig. 4.4.

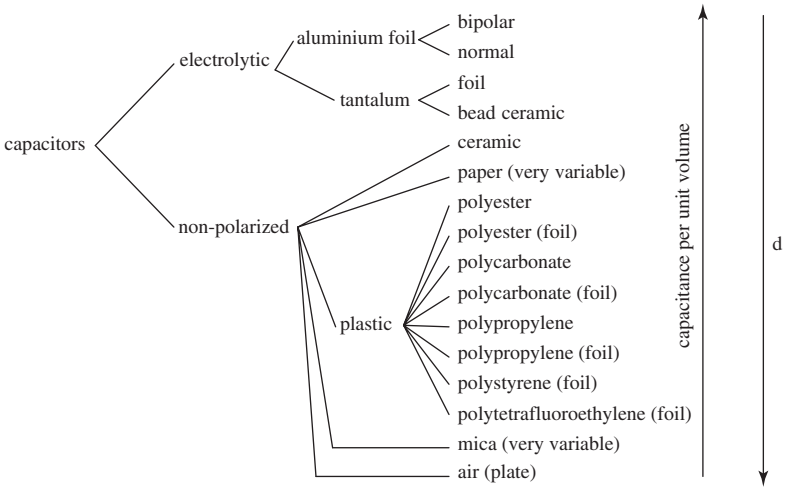


Fig. 4.4 *Comparison of different types of capacitor*

This tree of capacitors shows the various possibilities available. The first branching is between polarized and non-polarized capacitors. A polarized capacitor would be damaged by having DC applied in the reverse direction. Non-polarized capacitors branch into their plate construction; self-supporting plates, foil, or a surface coating of metal sputtered directly onto the dielectric. The final branchings deal with the dielectric, and although some dielectrics are represented in both categories, others are not, due to their manufacturing impossibility.

Broadly speaking, the more nearly perfect capacitors are at the bottom of the tree, whilst high capacitance per unit volume capacitors are at the top of the tree. This can be further generalized by observing that high quality capacitors tend to be physically large for their value of capacitance.

Air dielectric, metal plate ($\varepsilon_r \approx 1$)

These capacitors are invariably constructed as trimmer or variable capacitors with sets of intermeshing semicircular rigid plates, and are primarily used in radio frequency (RF) circuits, although they are occasionally useful in audio. Because of the difficulty of supporting plates that are very closely separated, air dielectric capacitors have low values of capacitance and are not usually larger than 500 pF. They generally have $\approx$10: 1 range between maximum value (vanes fully meshed) and minimum value (vanes fully separated).

Possible audio uses include:

- ≈300 pF variable capacitor across the input of a moving magnet RIAA stage to allow cartridges to be optimally loaded by the pre-amplifier.
- ≈50 pF for trimming equalization capacitors to their exact value.

Valve short wave radios often contain many smaller 'beehive' trimmer capacitors, and although they might not be the exact value required for your particular application, the silver-plated brass vanes are soldered to their supports, so they can easily be desoldered to reduce the value as necessary. See Fig. 4.5.

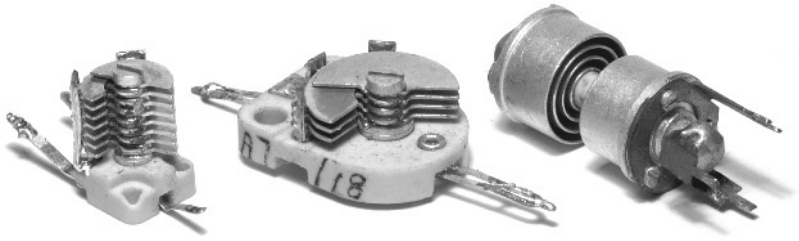


Fig. 4.5 *Selection of variable air-spaced capacitors. Note that the right-hand trimmer has its vanes disengaged so that they can be seen*

Plastic film, foil plate capacitors ($2 < \varepsilon_r < 4$)

This is the most important class of capacitors for use in valve amplifiers, as we will use these for coupling stages and also for precise filters. These capacitors are very nearly perfect, so their imperfections are usually described by the value of tan δ, or '*d*'. There appears to be a strong correlation between the subjective sound quality of capacitors and their value of '*d*', with low '*d*' capacitors being subjectively superior.

The significance of '*d*' in engineering terms is not simply that the capacitor has a leakage resistance across it, but that the capacitor is actually a ladder network of capacitors, separated by resistors, that extends indefinitely. See Fig. 4.6.

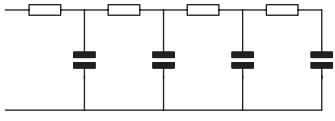


Fig. 4.6 *Equivalent circuit of practical capacitor to model dielectric absorption*

If we were to charge a capacitor whilst monitoring its terminal voltage with a voltmeter of infinite resistance, then discharge it by briefly short circuiting it, we would expect the capacitor voltage to remain at 0 V. However, we actually see the voltage rise from 0 V the instant that the short circuit is removed. This is because we discharged the capacitor that is 'near' to the terminals, but the other capacitors were isolated by series resistors, and were not discharged. Removing the short circuit allowed the undischarged capacitors to recharge the 'near' capacitor, and the voltage at the capacitor terminals rose. This effect is known as *dielectric absorption*, and it is more pronounced as the value of '*d*' rises.

Applying a pulse to a capacitor is equivalent to instantaneously charging and discharging the capacitor, so any voltage left on the capacitor at the end of the pulse is distortion. Music is made up of a series of transients, or pulses, and it may be that dielectric absorption is one cause of capacitor 'sound'.

Some plastics are *polar*, this does not mean that the capacitor can be damaged by reversing the polarity of any applied DC, but that at a molecular level within the dielectric, there are permanently charged electric dipoles, similar to the magnetic dipoles in a magnet. Under the influence of an external electric field, these dipoles attempt to align themselves to that electric field. When we apply an AC field, energy is absorbed as we successively align these dipoles first one way, and then the other, so that we incur a loss that rises with frequency. As frequency rises still further, we are no longer able to align the dipoles, and the loss begins to fall.

The dipoles may be imagined as individual blocks of identical mass, pivoted at their centre and resting on a rubber surface, that are rotated by pulling on rubber bands tied to their ends. As we pull on an individual rubber band, nothing happens until we overcome stiction (static friction), but once a block moves, it flips quickly to the new orientation because dynamic friction is always lower than stiction. Initially, as we pull on all of the rubber bands, only a few blocks rotate, because the stiction between individual blocks and the rubber surface is not constant. As we try to flip the dipoles faster and faster, we use more energy in overcoming stiction, but there comes a point when the applied frequency is so high that we simply vary the tension of the rubber bands and progressively fewer blocks move because they are decoupled. As fewer blocks move, we overcome less stiction, less energy is absorbed, and so the losses begin to fall.

This mechanism is so similar to magnetic hysteresis that it is sometimes known as electrostatic hysteresis, and the model is equally good for explaining magnetic losses.

Because of this frequency dependent loss, which begins to show at mid audio frequencies and peaks at low radio frequencies, capacitors using polar

plastics as their dielectric are not ideal for audio use. By comparison, non-polar dielectrics have losses that are independent of frequency almost up to UHF. Almost all dielectrics with $\varepsilon_r > 2.5$ are polar.

Dielectric	Common name	ε_r	*d*	Polar?
Polytetrafluoroethylene	PTFE, Teflon™	2.1	0.0002	N
Polystyrene		2.6	0.0002–0.0005	N
Polypropylene		2.2	0.0005	N
Polycarbonate		3.2–3.0	0.001–0.01	Y
Polyethyleneterephthalate	PET, polyester	3.2–3.9	0.002–0.015	Y

Film/foil capacitors are constructed by laying four alternate layers of dielectric and foil which are then wound into a cylinder. Guiding these four layers whilst winding the capacitor tightly is not a trivial task, and is partly responsible for the higher price of these capacitors. The foils are wound so that they are slightly displaced laterally relative to one another, so that one end of the cylinder has a spiral of foil that is one plate, whereas the spiral at the other end is the other plate. Each end is then sprayed with zinc to connect all points of the spiral together (which greatly reduces inductance) and leads are welded or soldered to this layer.

Because of the low melting point of polystyrene, traditional small (< 100 nF) polystyrene capacitors could only allow single point contact at $\approx 2/3$ from the centre of the winding, resulting in significant stray inductance. Nevertheless, some modern stacked foil polystyrene capacitors have circumvented the melting problem, and make contact over the entire foil. See Fig. 4.7.

Polystyrene capacitors often have one end of the capacitor delineated by a red or yellow band. This does not mean that they are sensitive to polarity, but that the marked end is connected to the outer foil. This is significant because one end of the capacitor may be connected to a less sensitive part of circuitry than the other. For instance, if a small polystyrene capacitor was used as part of an active crossover network, and connected as a series coupling capacitor (high-pass filter), then the banded end should be connected to the source to reduce induced hum. Alternatively, if one end were connected to ground, then the banded end should go to ground to reduce stray capacitance to active signals (strays to ground rarely cause problems, but Miller effect can cause other strays to be significant).

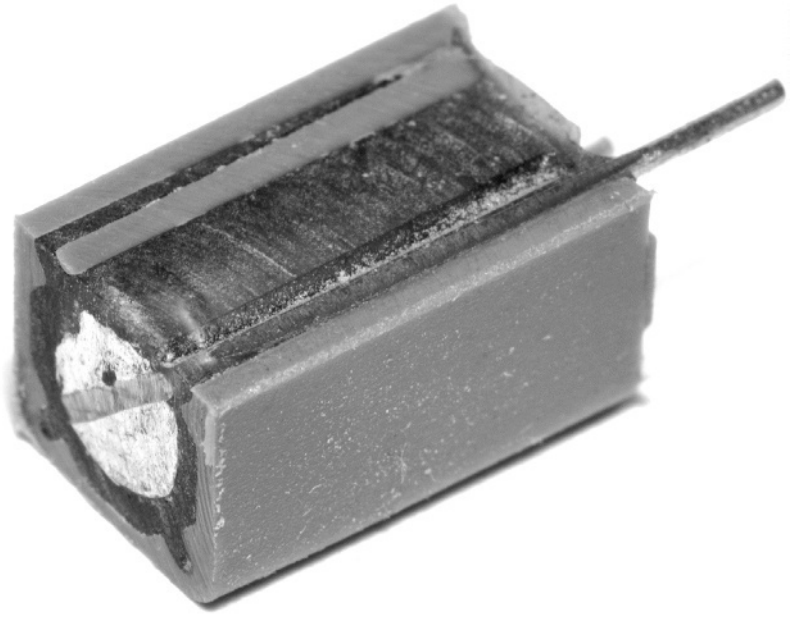


Fig. 4.7 *Dissected polystyrene capacitor*

Metallized plastic film capacitors

Because of the difficulty of winding the layers of dielectric *and* foil, most film capacitors are made by sputtering one side of the film with a layer of aluminium up to 12 μm thick to form the plate. Not only does this make the capacitor easier to make, but a far higher capacitance per unit volume is obtained because the plate is so much thinner, but ESR rises. Since ESR for a plastic capacitor is only significant at very high frequencies, when it becomes comparable with the reactance of the capacitor, this is not too much of a problem. However, *foil* capacitors are often described by their manufacturer as being more suitable for high frequency pulse applications for this reason.

If there is granularity due to impurities in the film of a metal film resistor, this generates excess noise, and film resistors are always noisier than wirewound resistors. Since the plates in a metallized film capacitor are also produced by sputtering, it is not unreasonable to suppose that they will suffer from the same quality control problems – with the difference that capacitors are not routinely tested for modulation noise. Subjectively, foil capacitors seem to sound better than metallized film, and it is possible that this is the reason.

Metallized paper capacitors ($1.8 < \varepsilon_r < 6$)

Metallized paper was the traditional dielectric for capacitors in classic valve amplifiers, and depending on the paper and its impregnant, the performance ranged from poor to quite good. Unfortunately, if the seals of the capacitor are less than perfect, humidity enters and the capacitor becomes electrically leaky. The author once bought a Leak Stereo 20 power amplifier having paper coupling capacitors, every one of which had gone leaky.

Oil or epoxy resin impregnation of the paper improves matters considerably, such that the resulting capacitor is almost as good as a polypropylene capacitor. Because paper capacitors are inherently self-healing, they are widely used in the power generating industry. In the event of an overvoltage spike, the paper breaks down at its weakest point and the metallization at that point is vaporized, thus preventing a short circuit and catastrophic failure.

Silvered mica capacitors (muscovite mica $\varepsilon_r = 7.0$)

This was the traditional small value capacitor used for RF circuitry, or for audio filters where excellent stability of value was important. Mica is a crystalline rock that can be easily cleaved into fine sheets, which are then coated with silver, and a stacked construction gives low inductance.

Since mica is a natural material, it is subject to all the accompanying vagaries of inconsistency. There are various types of mica, but muscovite makes the lowest loss capacitors. Although muscovite mica and polystyrene have comparable losses ($0.001 < d < 0.0002$), dielectric absorption is 80 times worse for muscovite mica than polystyrene,[2] so polystyrene is generally preferred.

Because of cost and variability, silvered mica capacitors have now been almost eclipsed by polystyrene, which in turn is being superseded by polypropylene.

Ceramic capacitors

These have no place in the path of analogue audio!

Until now, the dielectrics that we have seen have had $\varepsilon_r < 10$, but 'high k' ceramic capacitors can achieve ε_r (or k) $\approx 200\,000$! Commonly, ceramic capacitors are made of barium or strontium titanate, which are both piezoelectric materials. This means that they generate a voltage when mechanically stressed (these materials were the basis of ceramic cartridges used by inferior 'music centres' for playing vinyl records).

Ceramic capacitors excel as high frequency bypasses in digital or heater circuitry where their poor stability of value and low '*d*' are not a problem.

Electrolytic capacitors

These capacitors are polarized. Reverse biased, they form quite a good quality short circuit, and damage the driving circuitry, whilst the capacitor expires to the accompaniment of heat, smoke and noxious fumes. Aluminium electrolytics may even explode and shower surrounding circuitry with wet electrolyte and aluminium foil, causing further damage.

Some people have religious convictions against using electrolytic capacitors, but with all their faults, they are still useful components, and our choice of design is severely restricted if we refuse to use them. Most of the faults ascribed to electrolytics relate to inappropriate usage.

Electrolytic capacitors take high capacitance per unit volume to the limit, and they do this by attacking all parts of the parallel plate capacitor equation. The gap between the plates is minimized, surface area is maximized, and $\varepsilon_r \approx 8.5$ for aluminium oxide, as opposed to $\varepsilon_r \approx 3$ for the plastic films. The principle of operation is broadly similar for all types, so only the aluminium type will be described in detail.

Aluminium electrolytic capacitors ($\varepsilon_r \approx 8.5$)

The aluminium foil of one plate is anodized to form an insulating layer of aluminium oxide on the surface ($\approx$1.5 nm/V of applied polarizing voltage), and it is this micro-thin layer that is the dielectric of the capacitor. Since anodizing is an electrochemical process, and the aluminium oxide is an insulator, it follows that there must be a maximum thickness of aluminium oxide that can be produced before the insulation of the layer prevents deeper anodizing. This means that there is an absolute maximum voltage limit for electrolytic capacitors. Traditionally, reliable capacitors could only be made up to 450 V_{DC} working, but some modern capacitors are rated at 600 V. Old $>$ 450 V capacitors should be looked upon with grave suspicion.

Although by anodizing the aluminium foil, we have both a plate and the dielectric, we still need the other plate. We could use another piece of aluminium foil pressed tightly to the first, but any gap between the two foils would negate the advantage of the micro-thin dielectric. The second plate is therefore made of thin soggy paper or simply a gel, which because it is wet, makes perfect contact with the anodized surface of the first plate, and this is the electrolyte from which the component derives its name. The electrolyte is

not a particularly good conductor of electricity, and so a second aluminium foil is laid on top of the electrolyte to allow a low resistance plate to be made.

We now have two aluminium foils separated by electrolyte that can be rolled into a cylinder to make our capacitor. If, before anodizing the aluminium foil, we had etched the surface of the aluminium foil, this would roughen the surface, and would greatly increase the surface area on a microscopic level. Since the electrolyte plate is in perfect contact with this surface, we have dramatically increased the area of the plates of the capacitor, and the capacitance rises accordingly.

Unfortunately, the electrolytic capacitor does have its disadvantages. Electrolyte resistance is significant, so deep etching of the foil increases the resistance from the bulk of the foil to the extremities that form the plate, and we can expect the highest capacitance per unit volume components to have a higher ESR. Not only do these tortuous paths into the crooks and crannies increase resistance, but they have limited current carrying ability before heating significantly and causing the electrolyte to evaporate. Compact capacitors therefore not only have high ESR, but also a low ripple current rating.

However, Sanyo's 'OS-CON' range of capacitors uses an organic semiconductor electrolyte that significantly reduces ESR. Reduced electrolyte resistivity allows deeper etching, resulting in higher capacitance per unit volume, and therefore reduced inductance. OS-CONs therefore have measurably better high frequency characteristics, and would be ideal as cathode bypass capacitors if it were not for their extremely high price.

Historically, electrolytic capacitors had poor tolerance on their capacitance (typically $+100\% - 50\%$). Although modern types are typically $\pm 10\%$, we should *never* use an electrolytic capacitor in a position where its value could not be safely doubled or halved without upsetting operation of the circuit.

If we make contact to a foil at one point, perhaps at the one end, the capacitance at the far end of the foil is in series with the inductance of the foil. Spraying molten zinc onto the entire spiral connects all parts of each plate together, which is equivalent to an infinite number of connections, and minimizes inductance. This technique is not possible with an electrolytic capacitor, so connections to the plates have to be made with individual tabs of foil at various points along the spiral. Increasing the number of tabs reduces inductance, but they are difficult to make, and the greater the number of turns in the spiral, the more tabs are needed to keep the inductance down to a given value. Better quality capacitors reduce this problem by having a larger aspect ratio (can length divided by diameter) to obtain the required capacitance.

Although the manufacturer aims to minimize series inductance and thus inductive reactance ($X_L = 2\pi fL$), for a large capacitor, X_C is also low, so the inductance of an electrolytic capacitor is proportionately significant. Manufacturers' data sheets generally describe the problem by specifying the self-resonant frequency for each value of capacitor. In general, the higher capacitance types have a lower resonant frequency, which may be as low as tens of kHz.

Electrolytic capacitors are lossy. When they are first manufactured, a *polarizing* voltage is applied, and this causes a current to flow through the capacitor, forming the aluminium oxide layer on the plate. Once this oxide layer has been *formed*, very little current flows. However, this micro-thin layer gradually deteriorates with time and needs to be *re-formed*. Provided that the capacitor always has DC applied across it, the capacitor balances itself by always passing the minimum necessary current to maintain the oxide layer for the applied DC voltage.

If the equipment is switched off, when power is reapplied, a higher than normal leakage current flows until the oxide layer has been re-formed. The longer the elapsed time without bias volts applied, the greater duration and amplitude this initial leakage current will be, and there is a danger that this current will cause serious heating of the electrolyte. As the electrolyte is heated, it evaporates more readily, and the resulting gas may build up sufficient pressure to cause the can to explode. Because of this, it is wise to use a Variac to gently apply power to equipment containing electrolytic capacitors that has lain idle for some time.

Modern capacitors have safety pressure seals to vent the gas via a rubber bung in the base of the component (large capacitors), or the aluminium can may be deliberately weakened at the top with a series of indentations that allow controlled rupturing for the gas to escape (small capacitors). Either of these occurrences signifies the demise of the component, but they do prevent damage to other components, with the bonus of a simple visual inspection of their health.

Gentle heating evaporates the electrolyte through the seals of the capacitor (no seals can be perfect), and as the quantity of electrolyte falls, it makes less and less contact with the nooks and crannies of the etched plate, so the ESR rises and capacitance falls.

Electrolyte evaporation causes electrolytic capacitors to be heat sensitive, so capacitor life doubles for every 10 °C drop in temperature.

Applied voltage also affects electrolytic capacitor life. Without a bias voltage, the oxide layer cannot be re-formed and gradually deteriorates, causing the capacitor to become leaky. This was a well-known fault in

analogue sound mixers using symmetrical + and − supplies with operational amplifiers coupled by electrolytic capacitors, which then had little or no polarizing voltage.

Provided that there is a polarizing voltage, operating electrolytic capacitors below their maximum rated voltage increases their life significantly:

$$Life_{(\%\text{ of rated})} \approx \left(\frac{V_{\text{max.}}}{V_{\text{applied}}}\right)^5 \times 100\%$$

Using this relationship, we see that operating an electrolytic capacitor at 87% of its rated voltage doubles its life. It is wise not to read too much into this formula, since we could easily use it to predict a lifetime measured in centuries by lowering the operating voltage sufficiently. A good engineering rule of thumb is that, if possible, electrolytic capacitors should be operated at two thirds of their maximum voltage rating, giving a theoretical eight-fold increase in life expectancy, which is probably at the limit for which the formula is valid.

Many classic valve amplifiers have electrolytic capacitors with more than one component concentrically wound in a single can. The outer capacitor is marked with a red spot, and in an amplifier using cascaded RC smoothing, this capacitor should be connected to the most positive potential. The logic for this is that the highest potential has the greatest ripple voltage, and as there is no field within a conductor, this ripple is not coupled to subsequent stages. Wiring the capacitors in the wrong order causes excess hum.

There is a class of aluminium electrolytics available for use at AC that is known as *bipolar*. These capacitors used to be found in loudspeaker crossovers because they were usually cheaper than plastic film capacitors of comparable capacitance. Their construction is effectively two electrolytic capacitors back to back. See Fig. 4.8.

Fig. 4.8 *The bipolar electrolytic capacitor*

There is no constant polarizing voltage, and each individual capacitor has to be twice the value of the final capacitor. Defects are thereby multiplied by a factor of four over the normal unipolar electrolytic capacitor, so their performance is poor.

Tantalum electrolytic capacitors ($\varepsilon_r \approx 25$)

The increased relative permittivity significantly reduces component volume compared to aluminium electrolytic capacitors ($\varepsilon_r \approx 8.5$). Tantalum foil capacitors have two additional advantages directly related to the increased chemical inertness of the tantalum oxide layer. First, ESR can be reduced, by using a lower resistance electrolyte that would have corroded aluminium foil. Second, because the oxide layer is more inert, leakage current is reduced. However, tantalum is expensive, whereas aluminium electrolytic capacitors are improving all the time.

Tantalum bead capacitors are only available in low voltages, but the reduced inductance compared to an aluminium electrolytic capacitor makes them useful for bypassing voltage regulators or logic chips. Unfortunately, they are only available in low values (typically $< 100\,\mu F$), and are often not large enough to be used as cathode bypasses. When tantalum bead capacitors fail (they have zero tolerance to reverse polarity), they tend to fail dead short circuit, and the consequential damage can be spectacular. They are expensive, and becoming difficult to obtain.

General considerations in choosing capacitors

Voltage rating

- Will the voltage across the capacitor change polarity, or is it simply a varying DC? If the voltage across the capacitor is AC, then conventional electrolytic capacitors are eliminated.
- Will the capacitor withstand the applied DC voltage plus the maximum expected signal voltage (V_{pk}, not V_{RMS})?
- Could the capacitor withstand the maximum possible HT voltage? If not, what arrangements have been made to ensure that its rated voltage is never exceeded?

Tolerance

- Is the *absolute* value important? If it is part of a filter or equalization network, then we need a close tolerance component, such as polystyrene, polypropylene (now available in $\pm 1\%$), or perhaps silvered mica.
- Matching: is the capacitor part of a pair, such as coupling capacitors in a push–pull amplifier, or the corresponding component in the other stereo channel? If it is, then they should be matched if possible.

Heat

Will the capacitor become warm? Will the consequent change in value matter? In general, capacitors should not be operated at more than 50 °C (because insulation resistance falls with increasing temperature), but even this temperature could reduce the life of an electrolytic capacitor unacceptably.

Capacitance

Each type of capacitor is only available in a limited range of values, and if we need 330 μF, then only an electrolytic can provide this value at a sensible price and size.

Leakage and '*d*'

- Is leakage important? A cathode bypass or HT smoothing capacitor may be allowed to pass a small leakage current. A grid coupling capacitor may not be allowed to be leaky under any circumstances.
- Is this component important for final sound quality? Capacitors in the obvious signal path are important, but the signal *current* has to return through the HT supply, so HT smoothing and bypass capacitors are equally important. Smoothing capacitors for bias circuitry may be less important if there is no audio signal on them.

Microphony

All capacitors are microphonic to a greater or lesser extent. The reason for this is very simple. Suppose that we have stored a fixed charge on a capacitor:

$$Q = CV$$

The capacitance of a parallel plate capacitor is:

$$C = \frac{A \cdot \varepsilon_0 \cdot \varepsilon_r}{d}$$

Combining these equations, and solving for V:

$$V = \frac{Qd}{A \cdot \varepsilon_0 \cdot \varepsilon_r}$$

Since Q, A, ε_0 and ε_r are constants, if we vary the spacing between the plates, the voltage across the capacitor *must* change. This principle is the basis of the capacitor microphones used in studios and the ubiquitous electret microphone found in portable recorders.

The principle is reversible, and varying the applied voltage across a capacitor alters the attractive forces between the plates, and if they are free to move, this causes vibration. This is the basis of the electrostatic loudspeaker.

It might be thought that the plates of a plastic film capacitor would be sufficiently tightly wound that no movement was possible, but the author once built a stabilized HT power supply where the output bypass capacitor whistled loudly at ≈ 2 kHz. The circuit was thereby diagnosed as unstable even before the oscilloscope was ready!

The problem of capacitor microphony can be tackled in three ways, listed in descending order of desirability:

- Avoid using capacitors. To a limited extent, this is feasible.
- Isolate capacitors from vibration. Capacitors carrying low-level signals will be proportionately more sensitive to microphony than those carrying high-level signals. Pre-amplifiers are therefore most susceptible, and it is well worth isolating them from vibration. This is easily done at the design stage, but is much harder once built.
- Capacitors are physical objects, so they have mechanical or acoustical resonances. If we excite these resonances, we should expect them to be audible in the same way that striking a tuning fork produces an audible note. If we mechanically damp the capacitor by gluing it to another surface, then these resonances will be reduced. Provided that the capacitor can survive the heat, the soft glue used by hot-melt glue guns is ideal.

There is no reason why we should not use all three methods in combination if it seems that microphony is likely to be a problem. A good test for microphony is to tap each component with a plastic insulating pen (to avoid shock risk) with the equipment turned on whilst listening to the loudspeaker. The results may surprise you!

Bypassing

All capacitors become inductive as frequency rises, but if we bypass them with parallel capacitors of a smaller value, the composite component can more nearly approximate a perfect capacitor. Bypassing capacitors used to be very important, but modern electronics operates at higher frequencies, so capacitor design has had to improve, greatly reducing the need for bypassing.

A practical rule of thumb used to be that the ratio of main capacitor to bypass capacitor should be ≈100: 1. With very old components, more than one bypass might have been required. For instance, a 220 μF electrolytic capacitor would have been bypassed by a 2.2 μF plastic film capacitor, and perhaps also a 2.2 nF, but there was no point in carrying the principle any further because the inductance of the leads used to connect the capacitors together became significant. With modern capacitors, a single 100:1 bypass is quite sufficient.

Wires have inductance, and it should be realized that there is no point in having a beautifully bypassed capacitor if it is physically distanced from its point in the circuitry. Thus, power supply capacitors should be bypassed *at the load*, not at the main electrolytic capacitor. We cannot often connect all these capacitors directly between the output transformer HT tap and the cathode returns of the output valve(s), but we can, and should, connect the bypasses between these points. See Fig. 4.9.

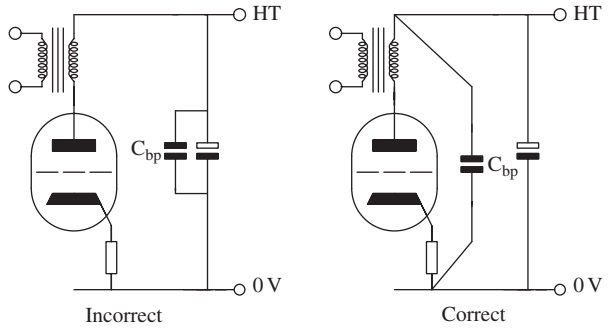


Fig. 4.9 *Connection of bypass capacitors*

A very useful assumption to make when considering the layout of *any* circuitry is to treat each wire as an aerial with inductance, and to assume that the circuit is immersed in a strong radio frequency (RF) electromagnetic field that induces currents into each and every wire. (This assumption is not so

very far from the truth.) It is for this reason that a bypass capacitor aiming to make the composite capacitor a perfect short circuit at RF should be connected to its point in the circuit by *very* short wires. Thus, the smallest bypass should be nearest to the load, whilst the larger capacitors may be further away. Each capacitor should be connected using individual wires to the load, so that they are 'starred' to the load point, this will minimize ground noise. If this results in seemingly untidy wiring, so be it.

Magnetic components

Magnetic components include transformers and inductors. Transformers may be signal transformers, such as output transformers and moving coil step-up transformers, or they may be mains transformers. Inductors may be the small-signal inductors used in filters, or they may be the power chokes in an HT supply.

Magnetic components are the least perfect *passive* components (resistors, capacitors and inductors/transformers), and for this reason, many designers shun them. This is unfortunate, because it seriously restricts design choice.

Inductors

Inductors store energy in the form of a magnetic field. Any wire passing current generates a magnetic field, so it must possess inductance. We can deliberately increase this inductance by winding the wire into a coil, whilst placing the coil around an iron core increases inductance still further. These proportionalities can be expressed by:

$$L \propto \frac{\mu_0 \mu_r A N^2}{l}$$

where: L = inductance
μ_0 = permeability of free space = $4\pi \times 10^{-7}$ H/m
μ_r = relative permeability of the core magnetic material
A = magnetic path cross-sectional area
l = magnetic path length
N = number of turns

Relative permeability is the magnetic analogue of relative permittivity that we met earlier, and has a value of 1 for air and ≈5500 for iron. The magnetic path length is the length through the core back to the starting point, and the cross-sectional area of the magnetic path is simply the cross-sectional area of the core, so it ought to be easy to derive a useful equation for calculating inductance.

Unfortunately, μ_r varies hugely with flux density; the path length is easily affected by air gaps; and some flux escapes the core. We will look at each of these problems in more detail in a moment, but suffice to say that we cannot often accurately calculate the inductance of a coil. We are forced to make an informed guess, add a few turns, measure the inductance under the actual operating conditions, and then remove turns until the desired inductance is achieved.

A curve that appears in virtually every discussion of magnetic materials is the *BH* curve. This is a plot of the relationship between applied magnetizing force and resulting magnetic flux. For our purposes, we need only note that μ_r is proportional to the gradient of the curve, and that as the gradient changes with level, so must μ_r. See Fig. 4.10.

Air-cored inductors

We can completely avoid the problem of non-constant μ_r with level by not using a magnetic material in the core. *Air-cored* inductors have constant

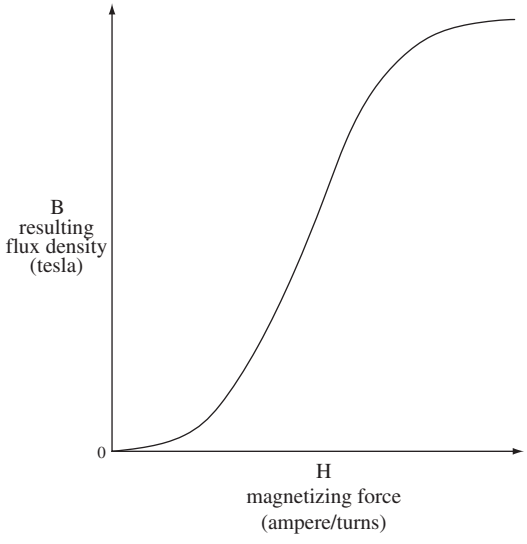


Fig. 4.10 *BH curve: non-constant slope implies non-constant* μ

inductance with applied signal level, and do not therefore cause distortion, making them popular in high quality loudspeaker crossover networks. Determining the magnetic path area is more difficult since this theoretically extends to infinity, whilst the path length is similarly awkward to define. Nevertheless, formulae have been produced for various core geometries, and a particularly useful set of formulae for optimum (lowest) resistance air-cored copper wire coils based on a paper by A. N. Thiele[3] is given below:

$$R = 8.01 \times 10^{-3} \sqrt[5]{\frac{L^3}{d^8}}$$

$$N = 10.2 \sqrt[5]{\frac{L^2}{d^2}}$$

$$c = \frac{d\sqrt{N}}{0.841}$$

$$l = 0.188\sqrt{Lc}$$

where: R = resistance in Ω
L = inductance in μH
d = diameter of wire in mm
N = number of turns
c = core radius (see Fig. 4.11)
l = length of wire in m

The formulae are given in this modified form because wire is available only in a range of standard diameters, and the resistance of the coil is not usually

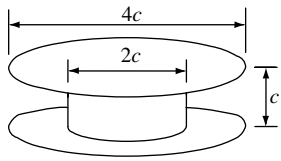


Fig. 4.11 *Relative bobbin dimensions for air-cored coils using Thiele formulae*

critical. If the resistance is different from that wanted, then a different wire diameter can be tried.

It is convenient to perform these calculations on a computer, so a QBASIC program is given below (although the key equations could equally well be dropped into a spreadsheet):

```
CLS
L = 1
PRINT ''This program designs air-cored copper''
PRINT ''wire coils according to the Thiele''
PRINT ''formulae. L is in microhenries, d (wire''
PRINT ''diameter) is in mm.''
PRINT
PRINT ''To quit, input L = 0.''
PRINT
DO WHILE L > 0
INPUT ''L''; L
IF L = 0 THEN END
INPUT ''d''; d
R = (((8.01 * 10 ^ (-3) * (L ^ 3 / d ^ 8) ^ (1 / 5)) *
100) \ 1) / 100
N = (((10.2 * (L ^ 2 / d ^ 2) ^ (1 / 5)) * 100) \ 1) /
100
c = (((d * N ^ (1 / 2) / 0.841) * 100) \ 1) / 100
Q = (((0.188 * (L * c) ^ (1 / 2)) * 100) \ 1) / 100
PRINT
PRINT ''You need''; N; ''turns on a core of''; 2 * c;
''mm in diameter,''; c; ''mm thick.''
PRINT ''It will use''; Q; ''metres of wire, and''
PRINT ''will have a resistance of''; R; ''Ohms.''
PRINT
LOOP
```

Experimentation soon reveals that air-cored inductors have significant resistance, or that they are very large. This problem of resistance is common to all inductors, and is one of their imperfections. Air-cored inductors are not only useful for loudspeaker crossovers, but also for the output filters of Digital to Analogue Converters (DACs), where the resistance is far less of a problem.

It should be noted that because of practical considerations (winding efficiency, variable wire diameter, etc.) the above formulae do *not* give exact

answers, and it is therefore wise to design 5% oversize and then remove turns whilst measuring the inductance with a component bridge.

Many component bridges use a 1 kHz internal oscillator. When measuring air-cored coils, the relatively high resistance may swamp the inductive component and the bridge will give misleading results. If it is possible to feed the bridge from an external source of AC, it should be fed with the highest frequency that the bridge manufacturer allows (typically 20 kHz), and this will allow sensible measurements to be taken.

Gapped cores

One way to reduce resistance without suffering gross distortion is to use a coil with a magnetic core that has an air gap. A *gapped* core significantly increases inductance over an air-cored coil, but because the air gap is effectively a large magnetic resistance in the path of the magnetic flux, it swamps the variation in permeability of the much smaller magnetic resistance of the core, and inductance is more nearly constant. As the gap becomes larger, inductance falls, and if the gap is infinitely large, then we are back to an air-cored coil. This technique was used for many years by the BBC Research Department for inductors in passive loudspeaker crossovers.

We may unintentionally make an inductor with a gapped core. Many ferrite cores used for small inductors are supplied as two mating halves that fit around the core once wound. Dust on the mating surfaces causes an air gap, and if the cores are gently squeezed together whilst inductance is measured, a significant increase in inductance may be seen.

If an inductor has to pass DC, it is essential that the DC current should not saturate the core, since this would drastically reduce the inductance. Iron-cored inductors passing DC are invariably gapped in order to maintain their inductance up to their rated maximum current, so they produce AC leakage flux near the gap that may interfere with nearby circuitry.

Self-capacitance

If a coil is made up of many turns of wire, and there is a potential between different turns and layers of turns, then we must expect the inductor to have capacitance in parallel with its inductance. See Fig. 4.12.

We now have our familiar parallel resonant circuit, which means that as we rise beyond the resonant frequency, the coil no longer looks like an inductor and behaves as a capacitor. The easiest way of finding this capacitance is to set up a test circuit. See Fig. 4.13.

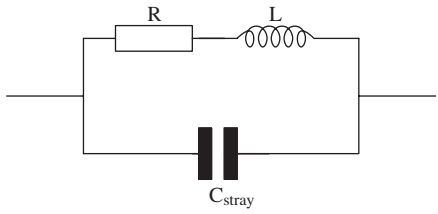


Fig. 4.12 *Equivalent circuit of practical inductor*

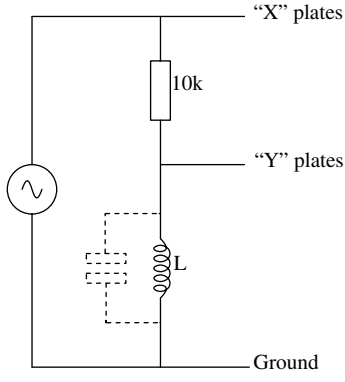


Fig. 4.13 *Using Lissajous figures to find self-resonant frequency of inductor*

The oscilloscope is used in XY mode and as the frequency of the oscillator is varied, the resulting Lissajous figure changes from an ellipse to a straight line; this is the resonant frequency of the inductor. If necessary, we can now calculate the shunt capacitance using:

$$C = \frac{1}{4\pi^2 f^2 L}$$

Power supply chokes for small valve amplifiers are typically 10–15 H @ 100–250 mA, and these chokes have a resonant frequency $\approx 3-12$ kHz. Above its resonant frequency, the choke cannot provide an effective barrier to the noise generated by rectification, or to RF incoming from the mains.

We will consider power supply chokes further in Chapter 5.

Transformers

In a perfect transformer, the magnetic flux of the primary winding is coupled to the secondary winding with no loss whatsoever. Practical transformers are somewhat different.

In a transformer, the losses are often divided into two distinct groups, *iron losses*, so called because they are due to imperfections of the core, and *copper losses*, due to imperfections of the windings.

Iron losses

As the core is successively magnetized and demagnetized through opposite polarities, work has to be done to change the alignment of the magnetic dipoles. This loss is known as *hysteresis* loss, and may be calculated by investigating the hysteresis curves for the particular core material used. Because it is the loss caused by changing the core magnetization through one complete cycle of the applied AC waveform, there will be more loss in a given time if more cycles of magnetization are traversed. Hysteresis loss is therefore directly proportional to frequency, and can only be reduced by choosing a lower loss core material.

Magnetic cores are metal, and therefore conduct electricity. As far as the primary winding is concerned, there is no distinction between an intentional secondary winding connected to a load and a conductive path parallel to the primary winding through the core. Conductive paths through the core cause *eddy* currents to flow, which, because they are short circuits, cause losses. To reduce these losses, the core can be constructed from a stack of *laminations* that have had their surfaces chemically treated to make them insulators. The ultimate approach to this problem is to make the core of iron dust particles whose surface has been treated, and then bond these with a ceramic to form a solid core known as a *ferrite dust* core.

Eddy current loss is proportional to f^2 because not only is the loss proportional to the number of traverses of the magnetization loop in a given time, but higher frequencies have smaller wavelengths and allow more loops of current to form within the core. Although thin steel laminations are satisfactory for audio frequencies, ferrites are necessary for RF frequencies, and at VHF, almost all core materials are excessively lossy, and air-cored transformers must be used.

Hysteresis and eddy current loss are often combined and known as magnetizing current in power transformers, and are responsible for core heating even when a load is not connected.

Not all of the flux from the primary flows through the secondary winding, and this loss, combined with hysteresis and eddy current loss, is known as *leakage inductance* in audio transformers. Theoretically, leakage inductance (referred to the primary) is found by measuring the primary inductance with the secondary short circuited. In practice, leakage inductance is difficult to measure accurately because a measurement at only one frequency is upset by stray capacitances. Nevertheless, leakage inductance is an important theoretical concept, since it determines the high frequency operating limit of the transformer.

Leakage inductance is dependent on the size (q), the turns ratio N^2, and the geometry of the transformer (k), but is independent of μ_r:

$$L_{\text{leakage}} \propto qN^2k$$

For a given frequency, a higher power rating transformer will be larger than a lower power rating transformer, and will consequently have higher leakage inductance.

Since leakage inductance is proportional to N^2, we should always try to keep the turns ratio as low as possible, so paralleling output valves in a valve amplifier is beneficial because it reduces the turns ratio required.

Geometry can be improved in two fundamental ways, we can either improve the shape of the core, or improve our winding technique.

Standard transformers are made with E/I cores, where each lamination of the core is composed of an E shape and an I shape. On alternate laminations, the orientation of the shapes is often reversed to reduce the air gap at the joint. See Fig. 4.14.

Traditionally, superior cores were made as C cores. These were made by winding the core out of a continuous strip, which was then cut in half, and the resulting faces ground smooth. The coils were then wound, and the cores were inserted so that the ground faces were perfectly aligned with minimal air gap, and steel straps were used to hold the assembly firmly together. See Fig. 4.15.

The C core was an expensive process, and inaccurate assembly could create an air gap, thus creating the very imperfection that the design intended to avoid. The more modern approach is to wind the core as a toroid, but not cut it, and use a special coil winding machine to wind the coils directly onto the core, resulting in a very low leakage core. See Fig. 4.16.

Incidentally, although toroids are thought of as being modern, the first transformer ever made was a toroid! (Michael Faraday, August 1831.)

Both the C core and the toroid have the further advantage that the magnetic flux always flows in the same direction relative to the grain direction

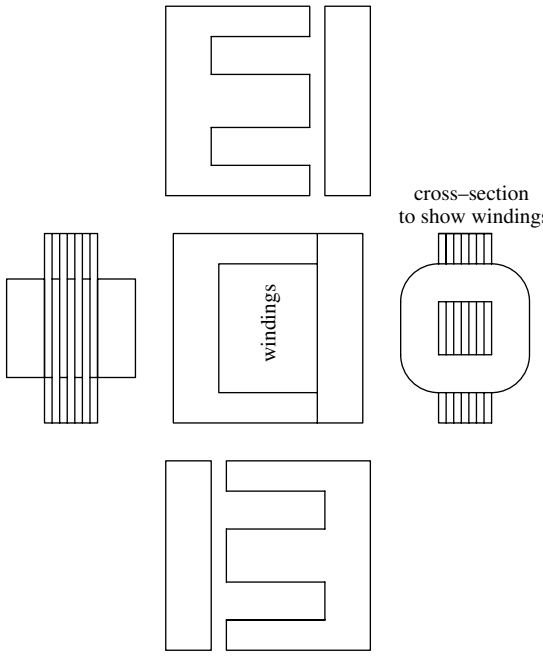


Fig. 4.14 *E/I core laminations arrangement to reduce leakage flux*

of the crystal structure of the core, whereas in the E/I core it has to flow across the grain in some parts of the core. This is significant because Grain Oriented Silicon Steel (GOSS) can tolerate a higher flux density before saturation in the direction of the grain, than across the grain. E/I cores can therefore only operate at flux densities below saturation across the grain, whereas C cores and toroids can operate at significantly higher flux densities, allowing core size to be reduced.

The geometry of the coil can be improved by winding the primary and secondary out of many interleaving layers or sections, rather than winding one

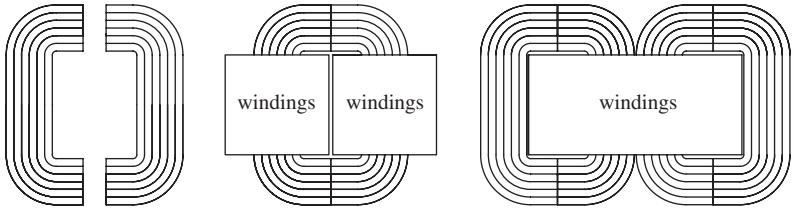


Fig. 4.15 *C core arrangement*

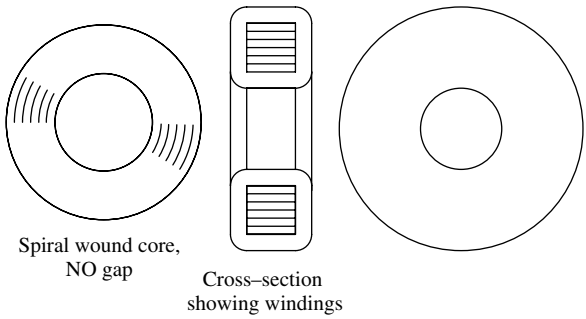


Fig. 4.16 *Toroidal core arrangement*

half of the bobbin with the primary and the other half with secondary. Increasing the number of sections improves the coupling between primary and secondary, thus reducing L_{leakage}, but probably increases stray capacitance. See Fig. 4.17.

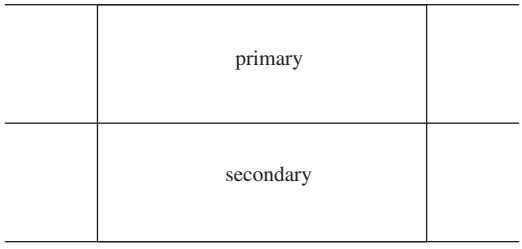


Fig. 4.17 *Winding geometry giving good primary/secondary isolation but poor leakage inductance*

Although sectioning the windings is relatively easy on an E/I or C core, it is very difficult on a toroid, moreover, winding geometry on a toroid is quite poor, and so it is easy to lose the benefits of the improved core by having a poor coil. Toroidal mains transformers are notorious for their leakage flux at the point where the windings exit for this very reason.

An alternative technique for improving winding geometry is to use *bifilar* winding, whereby two wires are simultaneously wound side by side. If one of these wires is part of the primary, and the other is part of the secondary, this promotes excellent coupling between the windings, and leakage inductance is significantly reduced. The technique is cheaper than sectioning, and providing the coil winding machine can cope, there is no reason to stop with two wires – three or four could be used.

Unfortunately, there are two snags to multifilar winding. First, the thin polyurethane insulation on the copper wire is easily damaged during winding, and may break down if we have >100 V between the windings, making it difficult to make a transformer capable of isolating the HT supply. Nevertheless, the seminal 50 W McIntosh[4] amplifier used a multifilar output transformer and a 440 V HT supply! Second, the greatly increased capacitance between primary and secondary may resonate with the reduced leakage inductance to produce a lower resonant frequency than a sectioned transformer.

Multifilar winding is best used in small-signal transformers with a very low turns ratio (ideally 1:1), such as the balanced line output transformers used in studios.

DC magnetization

If a net DC current is allowed to flow in a transformer, it shifts the AC operating point on the *BH* curve and causes significant distortion due to saturation on one half cycle. For this reason, output valve anode currents in push–pull amplifiers should be carefully balanced, and half-wave rectification should never be used on mains transformers. A traditional way of checking output valve DC balance in push–pull amplifiers is to measure the voltage between the anodes of the output valves, and adjust for zero volts. Zero voltage between anodes means equal voltage drops, and this implies equal currents with no out-of-balance current, *if the winding resistances are equal*. Checking these resistances before using this method is therefore essential. Because toroidal and R-type transformers do not have a gap, they are far more susceptible to core saturation due to DC.

Copper losses

Copper wire has resistance, and in a well-designed transformer, the losses due to resistance are equal in primary and secondary, and will therefore be related by:

$$R_s = \frac{R_p}{N^2}$$

where N is the primary to secondary turns ratio.

The total copper losses may in turn be traded against iron losses in a given transformer design, so that two transformers may have different proportions of iron to copper to achieve equal power ratings.

Electrostatic screens

Capacitance between primary and secondary sections is significant in audio transformers because it is multiplied by the turns ratio of the sections concerned, in a manner similar to Miller effect in the triode valve. The problem can be solved by interposing an earthed electrostatic screen, usually made of foil, between the affected windings. We now have capacitance to earth, but the effect of this capacitance is minimal. It is most important that the two ends of the foil do not contact electrically, as this would form a shorted turn.

An electrostatic screen between primary and secondary is often fitted to mains transformers for rather different reasons. If the insulation was to break down between primary and secondary, mains voltage would be connected to the secondary, which would be a safety hazard. By interposing an electrostatic screen, the fault current flows directly to earth and blows the mains fuse, thus making the equipment safe.

An electrostatic screen prevents RF interference on the mains from being capacitively coupled to the following circuitry. In audio, the significance of RF interference cannot be overemphasized, and this is sufficient incentive for using an electrostatic screen. Electrostatic screens are particularly beneficial for low voltage secondaries because they prevent high voltage noise from the mains being capacitively coupled directly into sensitive circuitry.

Magnetostriction

Valve amplifiers with output transformers tend to 'sing' audibly when operated at high power. Occasionally, this is due to a loose lamination, but it is more likely to be due to *magnetostriction*, which is an effect whereby a magnetic material changes its length according to the strength of the magnetic field passing through it. Output transformers support quite strong magnetic fields, so the effect can become noticeable. Since the magnetic field is varying, it causes vibration, but because magnetostriction is not polarity sensitive, in a push–pull amplifier, the sound that will be heard is pure 2nd harmonic distortion.

Magnetostriction is inversely proportional to μ_r, so a higher quality transformer is less likely to suffer from this (admittedly minor) problem.[5]

Output transformers, feedback and loudspeakers

It is far more convenient to derive feedback from a dedicated feedback winding, or from the end of a tapped winding, because it means that the

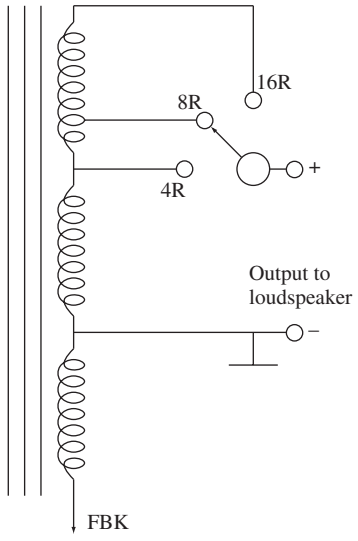


Fig. 4.18 *Secondary arrangement in Leak output transformer*

user can change the matching of the amplifier to the loudspeaker without having to adjust feedback. The Leak amplifiers were designed using this scheme, allowing a simple link to determine matching, but it means that the output transformer is not used optimally. See Fig. 4.18.

As an example, if the 4 Ω setting is chosen, then only half of the secondary winding is used, resulting in poorer leakage inductance. Worse, the feedback (which would ideally be applied at the output terminals) has to be coupled via the tapped secondary before being applied, and the coupling from one part of the secondary to another cannot be perfect. The optimum way to apply feedback is to derive it from the amplifier output terminals (or even better, the loudspeaker terminals). Ideally, transformer performance should be optimized by using as many of the secondary sections as possible in a carefully controlled way.

Old transformers tend to have a pair of secondary sections which are connected in series for 15 Ω loudspeakers, and in parallel for 4 Ω. The sections are not necessarily from the same layers, so the sections have differing resistances and leakage inductances. Connected in series (15 Ω matching) this is not a problem, but when in parallel, the mismatched Thévenin sources drive current into one another, which is not ideal. Better quality transformers have four secondary sections to cleanly give 1 Ω, 4 Ω and 16 Ω, but still have the same problem as before if configured for 8 Ω.

Why not 16 Ω? 16 Ω loudspeakers would not need a particularly low source resistance for optimum damping and would be less upset by loudspeaker cable resistance. In addition, transistor amplifiers could be designed more easily, valve amplifiers could have their secondary sections optimized, and the reduced turns ratio would further improve the transformer. However, any manufacturer who introduces a 16 Ω loudspeaker would have that loudspeaker branded as inefficient, because a given voltage would produce 3 dB less acoustic power than the 8 Ω design – read published comments about the BBC LS3/5a (12 Ω). So, we are stuck with 8 Ω, but the trend is firmly towards 4 Ω.

Because modern loudspeakers are nominally 8 Ω (often with dips that fall lower), or a nearly resistive 4 Ω, it is better to treat all loudspeakers as 4 Ω– the slight loss of 8 Ω measured power is insignificant, but the boost in quality is worthwhile.

Transformer models

Because real transformers are such complex devices, it is usual to devise simplified models that attempt to represent operation at low, mid and high frequencies.

At low frequencies, the transformer may be represented as a perfect transformer in parallel with the primary inductance of the real transformer, driven by the non-zero resistance of the source. See Fig. 4.19.

The combination of source resistance and finite primary inductance creates a high-pass filter whose cut-off frequency is given by:

$$f_{-3\,\mathrm{dB}} = \frac{r_\mathrm{s}}{2\pi L_\mathrm{p}}$$

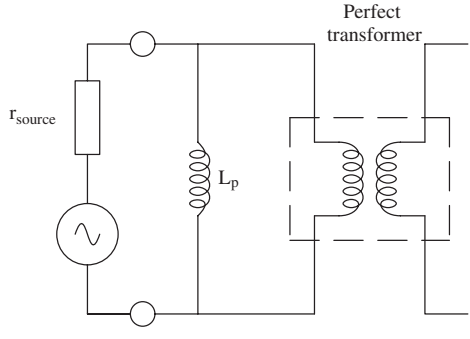


Fig. 4.19 *Transformer equivalent circuit at LF, showing effect of primary inductance*

With a given transformer, we will obtain better low frequency performance if we can reduce the source resistance. An EL34 pentode has $r_a = 15\,k\Omega$, but the same EL34, used as a triode, has $r_a = 910\,\Omega$, and used as a cathode follower, $r_k < 100\,\Omega$.

Unfortunately, the previous model is only appropriate for small signals. In a power amplifier, output valve operating conditions are invariably carefully matched to their load impedance, and the reduced reactance of L_p at low frequencies diverts signal current from the load into L_p. At high levels, the increased signal current flowing into L_p saturates the core, reducing L_p, further increasing signal current into L_p, leaving little current available for the loudspeaker load, so low frequency distortion increases catastrophically. Power amplifiers thus require the $f_{-3\,dB}$ frequency L_p to be determined by R_L rather than r_a, and the advantage of an EL34 cathode follower would not be seen at full power.

Once we have decided on the relevant resistance, we need a high primary inductance, which can be achieved either by increasing primary turns, or by using a core material with a higher μ_r. Although low frequency performance can obviously be improved by increasing μ_r, we could use increased μ_r to improve high frequency performance. Primary inductance would be maintained by winding fewer turns, thus reducing leakage inductance and stray capacitance, which would result in better high frequency performance.

A better core material is preferable because the bandwidth (in octaves) of a transformer with matched source and load resistances is:

$$BW_{(\text{octaves})} \approx \frac{\log\left(\frac{L_{\text{primary}}}{L_{\text{leakage}}}\right)}{\log 2}$$

The bandwidth is dependent on the geometry of the transformer, and on μ_r, but not size or number of turns. All other things being equal, a core with higher μ_r produces a transformer of greater bandwidth, which could be achieved either by a better core material, or by eliminating air gaps (toroid), or both.

At mid frequencies, we can consider the losses due to the resistance of the windings; it is usual to reflect the secondary circuit into the primary circuit. See Fig. 4.20.

The high frequency model is much more complex. See Fig. 4.21.

In this model, the primary circuit has been reflected into the secondary, and the source resistance, primary resistance, and secondary resistance have

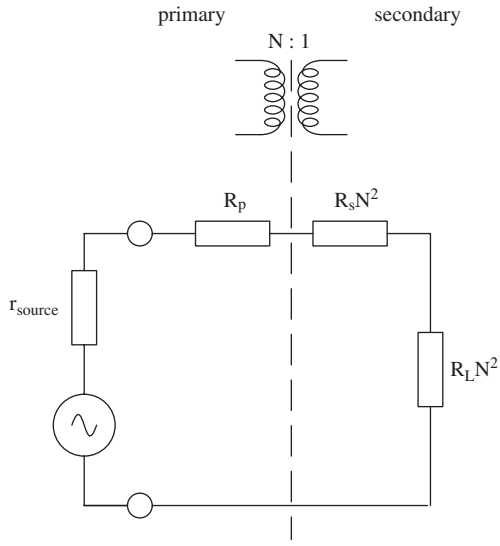


Fig. 4.20 *Transformer equivalent circuit at MF, showing effect of winding resistance*

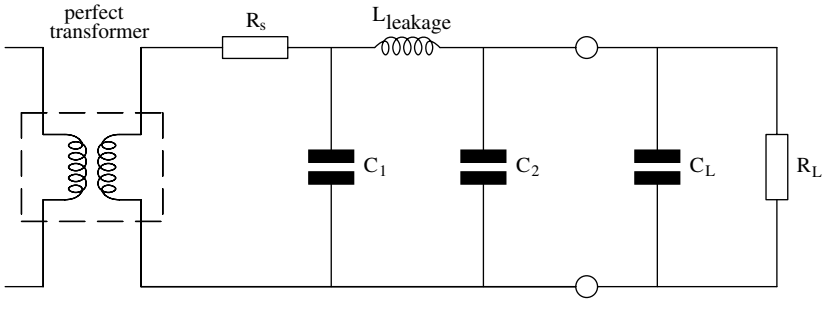


Fig. 4.21 *Transformer equivalent circuit at HF, and its similarity to classical 3rd order filter*

been lumped together. Interwinding capacitance has been lumped and introduced in two positions, and leakage inductance is also included. The resulting circuit is a classic low-pass filter having an ultimate roll-off of 18 dB/octave, and with suitable choice of component values, this model accurately simulates a real transformer at high frequencies.

Since the model is a classic filter, we can use the rules that apply to these filters. The most important of these rules is that performance is critically dependent on terminating resistances. For a normal filter, these terminating resistances are the source and load resistance, but for the transformer, we should also consider load capacitance.

Input transformer loading

When using a moving coil cartridge step-up transformer, it is well worthwhile experimenting to find the optimum load resistance for the *transformer*, before adjusting the load on the cartridge.

Once the source resistance to the transformer is known, the optimum loading for the input transformer can be determined. Beware that better cartridges tend to have higher coil resistances (because thinner wire reduces moving mass), so upgrading a cartridge could require a change of transformer loading. Not only will the frequency response be affected, but a higher cartridge resistance could cause significant loss in the inevitable potential divider formed by the cartridge resistance and the reflected transformer loading resistance.

As an example, the Sowter 8055 was originally designed for a 3 Ω cartridge, and its optimum loading resistance was then a pure 2.7 kΩ resistance. Since the step-up ratio is 1:10, and impedances are transformed by a ratio of n^2, the 3 Ω cartridge saw a reflected resistance of 27 Ω, giving a loss of 0.9 dB. Replacing the 3 Ω cartridge with a 10 Ω model increases the loss to 2.7 dB, so a further 1.8 dB of sensitivity has been lost.

The significance of the additional 1.8 dB loss is that the noise at the input of the amplifier has remained constant, so the change of cartridge source resistance has caused a loss of 1.8 dB in S/N ratio. Coincidentally, increasing the *gm* of the input valve by 50% gives an improvement of 1.8 dB in S/N ratio, but increasing *gm* is always fearfully expensive, so we must avoid unnecessary losses before amplification. If we could increase the loading resistance on the transformer, the reflected resistance seen by the cartridge would rise, and the S/N ratio would improve.

Unfortunately, any transformer has a high frequency resonance caused by leakage inductance and interwinding capacitances, and increasing the loading resistance reduces damping, producing a peak in the frequency response and ringing on square waves. However, a carefully chosen Zobel network across the secondary can significantly tame the ringing. Values for the network are easily found by experiment. See Fig. 4.22.

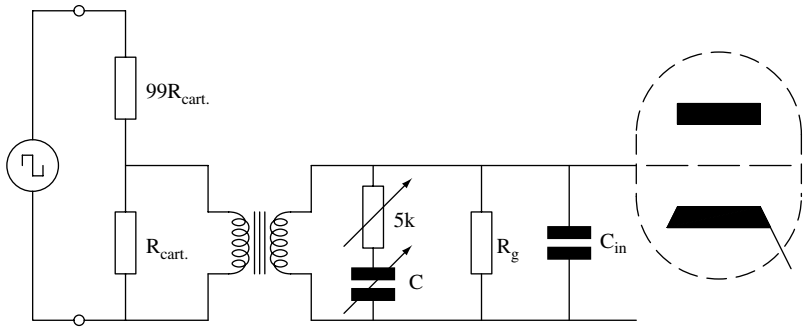


Fig. 4.22 *Determining Zobel network values for moving coil transformer*

The purpose of the potential divider at the output of the square wave generator is two-fold:

- We want to drive the transformer from the same resistance as the cartridge plus arm wiring resistance. Typical generators do not have 10 Ω output resistance, so the potential divider is designed to have the correct output resistance.
- The output voltage of typical square wave generators is far too high for the transformer, so we can easily afford to attenuate by a factor of 100.

Calculating the potential divider rigorously is unnecessary because physical coil winding constraints mean that cartridge coil resistances cannot be equal (5% error is typical). Additionally, because the potential divider needs to attenuate by a factor of 100, $r_{out} \approx R_{lower}$, so we just set R_{lower} to be equal to the required resistance, and choose the nearest convenient value to make $R_{upper} \approx 100R_{lower}$.

We know that the transformer will drive a valve that has input capacitance, so this should be calculated, or measured using the method given later in this chapter. Although ×10 oscilloscope probes reduce capacitance at the probe tip, they do not eliminate it, so this capacitance must also be considered during measurement. In this example, the transformer must drive an EC8010 triode whose input capacitance has been measured to be 190 pF. The author's Tektronix P6139A ×10 probes present a tip capacitance of 8 pF, so a 180 pF loading capacitor was used (180 pF + 8 pF ≈ 190 pF).

The Zobel resistance is unlikely to be larger than the main loading resistor, so a 5 kΩ linear potentiometer was used.

The variable capacitance in the Zobel network was salvaged from an irredeemably faulty valve AM radio (these used to be plentiful, but you might now need to go to a radio fair). Air-spaced variable capacitors typically achieve 300 to 500 pF with the vanes fully closed, but sections can be parallelled if necessary.

Once the generator has been set to produce a 1 kHz square wave with an amplitude of $\approx 100\,mV_{pk-pk}$ at the output of the transformer, the Zobel resistance and capacitance can be simultaneously adjusted to give the best possible leading edge as viewed on the oscilloscope. In general, the resistor adjusts how much curve there is at the leading edge, whilst the capacitor adjusts the ringing amplitude. Finding the optimum point is surprisingly easy.

Once the optimum resistance and capacitance have been set, they can be carefully desoldered and measured. Your DVM may claim to be able to make both measurements, but a component bridge will almost certainly be better for the capacitance measurement.

Why should I use a transformer?

With all that has been said about the imperfections of magnetic components, it might be thought that they should be avoided at all costs, particularly since they are invariably expensive.

An output transformer can be used to match a low impedance loudspeaker to the high resistance valve output stage, thereby greatly increasing efficiency. If multiple secondary windings are provided, it also allows user-selectable matching to various impedances without having to redesign the amplifier.

An input transformer, such as a moving coil cartridge step-up transformer, can step-up a small signal sufficiently so that it can be amplified by the following amplifier with minimum noise due to the amplifier. As a bonus, the primary can be left floating so that any hum induced into the connecting cable from the cartridge to the transformer is rejected by the transformer. (See Chapter 7 for a fuller explanation of these benefits.)

With the possibility of multiple windings, a transformer may allow novel methods of feedback to be applied to a circuit, further improving its performance. This technique has frequently been exploited in power amplifiers.[4]

Arguing the case for interstage transformers is rather harder. They invariably have to match high source and load impedances, requiring large inductances causing large stray capacitances that reduce bandwidth. Nevertheless, when expense is a secondary consideration, output valves such as the

845 may benefit from being driven by a robust driver valve coupled via a carefully specified and designed driver transformer.

A transformer isolates the DC on the primary from the DC on the secondary. This is often essential!

General considerations in choosing transformers

These considerations only apply to audio transformers; power supply transformers will be considered in the next chapter.

Unless you are building a standard circuit, for which a transformer has already been designed, you will almost certainly need a custom-designed transformer. It is therefore essential that you give the designer as many clues as possible so they may make the *right* compromises to suit your circuit:

- Is the transformer a power output transformer, or is it a small-signal transformer?
- What is the maximum signal level (mV) that will be applied to the primary at the lowest frequency of interest? Is this level constant with frequency? How much distortion can you tolerate at this frequency/level?
- What is the source resistance?
- What primary to secondary turns ratio is required?
- What shunt resistance and shunt capacitance will load the secondary? Can either of these be varied if necessary?
- What frequency range do you *need* the transformer to cover? Don't just say 5 Hz – 500 kHz ± 0.1 dB, because it can't be done.
- Do you need an electrostatic screen?
- Do you want the transformer screened in a μ-metal can to reduce electromagnetic hum?
- Are there any special requirements that the designer ought to know about?

If the answer to the first question was 'power output transformer', then these additional questions should be answered, and an annotated circuit diagram of the output stage is ideal:

- Is the output stage Class A or Class AB?
- What is the quiescent DC current? What is the maximum DC current?

- What is the maximum output power, and what is the lowest frequency at which this is required, for a given distortion level?
- Is the output stage push–pull, or single ended?
- Are the output valves triodes or pentodes? Will you need 'ultralinear' taps?
- What primaries and secondaries do you need? What DC is superimposed on each?
- What form of physical mounting do you want? Open flanges, shrouds, or drop-through?

All these questions may seem rather off-putting, but if you already have a clear idea of what you want, it is much more likely that the finished article will meet your expectations.

Identifying an unknown output transformer

Occasionally, you may find an old output transformer that appears to be of good quality, but you have no information as to what it is. Fortunately, identifying old output transformers using only a DVM is quite easy, as they usually conform to a pattern.

Before testing, all external links on the transformer should be noted and removed. (A digital camera is extremely useful here.) The primary is invariably centre tapped to allow for push–pull operation, and there may be further tappings to allow for ultralinear operation. Typically, the DC resistance measured by a multimeter from end to end will be the maximum winding resistance that can be measured, and is commonly between 100 Ω and 300 Ω. Once you have found a resistance of this order, you have identified the A_1 and A_2 connections.

Best quality transformers have the primary wound symmetrically, with equal resistance from the HT centre tap to both A_1 and A_2, so the next step is to find the connection that has half the end to end resistance. However, cheaper transformers may not be wound quite so well, and may not have perfectly equal resistances.

Because the same gauge wire is invariably used for the entire primary of the transformer, a tap that is 20% (maximum power configuration) from the HT centre tap to A_1 or A_2 has 20% of the resistance from HT to A_1 or A_2. If it is from a higher quality amplifier, then 47% taps (minimum distortion configuration) are more likely.

The secondary is likely to be composed either of an even number of windings, or one tapped winding. Remember that in the valve heyday,

loudspeakers were either 15 Ω (best quality loudspeakers), or 4 Ω, so output transformers were optimized for these impedances.

The most common configuration is to have two identical sections wired in series for 15 Ω, or in parallel for 4 Ω (actually 3.75 Ω). If, having identified the primary, you find two windings with ≤0.7 Ω DC resistance each, then you have a standard pattern transformer.

Better quality transformers took the above idea rather further, and had four equal sections comprising their secondary. Wired in series, they again matched 15 Ω, but with all four in parallel, they matched 1Ω. This was not because loudspeakers with 1Ω impedance were available (the technology for designing bad crossovers had not yet arrived), but because further sectioning of the transformer enabled a better quality component. This time, you are looking for four roughly equal DC resistances of ≈0.3 Ω. Bear in mind that contact resistance is very significant when making these measurements, so not only do you need a clean, firm contact, but also that even a 4½ digit DVM is inaccurate when measuring such low resistances, so a certain amount of guesswork is involved.

If after identifying the primary, all of the remaining wires appear to be connected together, then you have a tapped secondary, and the connections with the most resistance are between the 0 Ω and (probably) the 16 Ω tap. Provided that an 8 Ω tap is not fitted, the smallest DC resistance from *either* of these is the 4 Ω tap, and the 0 Ω is nearest to the 4 Ω tap (tapped secondaries tend to use thicker wire for the 4 Ω tapping). If an 8 Ω tap is suspected, then the tappings are best found using the AC method to be described later.

If some windings are still unaccounted for, these are likely to be feedback windings, perhaps to individual output valve cathodes, or for global feedback. Either way, these can be identified later, as the next step is to determine the turns ratio, and from this, the primary impedance of the transformer.

Although (when correctly carried out) the following test is not likely to be dangerous to the output transformer, it does expose LETHAL voltages, and if you are in ANY doubt as to your competence to perform the test, it should not be attempted.

Valve output transformers are designed to step down signals at hundreds of volts down to tens of volts at frequencies from 20 Hz to 20 kHz, so applying mains voltage from A_1 to A_2 is well within their capabilities. Providing that A_1 and A_2 have been correctly identified, there is no danger to the transformer in connecting the AC mains directly between A_1 and A_2, and measuring the AC voltage that appears across a secondary to determine the turns ratio.

The test should be done in the following manner:

- Fit a mains plug with the lowest value fuse you can find, 3 A will do, but 1 A is preferable.
- Wire the plug with a short length of three core flex. For obvious reasons, these leads are known as 'suicide leads', and should be kept locked away when not in use.
- Solder a solder tag to the 'earth' wire, and screw this to the metallic chassis of the transformer, remembering to use star washers to ensure good electrical continuity.
- Solder the 'line' wire to A_1, and the 'neutral' to A_2.
- Ensure that all links across secondary sections have been noted and removed.
- Set your DVM to its 'AC volts' range and connect it across one of the secondary sections.
- Ensure that you can see the meter clearly, plug the mains plug in, and switch on. If the meter does not react *instantly*, switch off. Assuming that a plausible reading appears, allow the reading to stabilize, note the reading, switch off, and unplug the mains plug.
- Check the mains voltage by reconnecting your DVM across A_1 and A_2, and applying mains again. Note the reading.

We can now find the turns ratio using the following formula:

$$N = \frac{\text{mains voltage}}{\text{secondary voltage}}$$

This may not seem particularly useful, but remember that impedances are related by N^2, so once we know N, we can determine primary impedance, since we already know the secondary impedance.

Complete example

Out of all its wires, the transformer has five wires that are all electrically connected when tested with the DVM on its resistance range. The largest resistance between two wires is 236 Ω, so these wires are designated A_1 and A_2. Leaving one DVM lead on A_1, another wire is found to give a resistance of 110 Ω, this is near enough to 118 Ω to be the centre tap, so this is designated HT. One DVM lead is now moved to the HT tap, and the two remaining primary wires are tested, one reads 29 Ω and the other reads 32 Ω, 29 Ω/110 Ω = 0.26, and 0.27 so these are likely to be maximum power

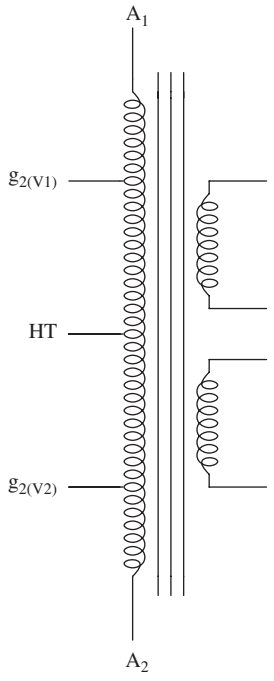


Fig. 4.23 *Identifying an unknown output transformer*

(≈20%) ultralinear taps. The one that reads the lowest resistance to A_1 is the $g_{2(V_1)}$ tap, and the other is the $g_{2(V_2)}$ tap. See Fig. 4.23.

The secondary has only two sections, so each is likely to be a 4 Ω section. This is then confirmed, as one measures 0.6 Ω, and the other 0.8 Ω, which are typical values for 4 Ω windings.

When connected, mains voltage read 252 V_{AC}, and the voltage across a single secondary was 5.60 V_{AC}. Substituting the values into the equation:

$$N = \frac{\text{mains voltage}}{\text{secondary voltage}} = \frac{252}{5.60} = 45$$

Impedances are changed by N^2, so the ratio of primary to secondary impedance is $45^2 = 2025$, and because the secondary voltage was measured across a 4 Ω section, the primary impedance must be $2025 \times 4 = 8100\,\Omega$. It is possible that testing with 252 V_{RMS} at 50 Hz might have driven the transformer close to saturation, causing errors, so we would round this value to 8 $k\Omega_{(a-a)}$.

We now need to 'phase' the secondary sections of the transformer, this is done by connecting one wire only between one section and another, to put the windings in series. If when AC mains is applied across the primary we measure double the original secondary voltage, then the two sections are aiding one another, and we have connected a 'phase' to a 'neutral', so we can designate the sections at the end of the linking wire '+', and the other '−'. However, if no voltage is seen, then we have connected like to like, so we can designate both as '+' or '−'.

If all identical sections have been identified and phased, any remaining windings can be measured and their turns ratio determined, either with reference to the secondaries, or to the primary – whichever seems the most convenient. At this point, annotated diagrams are usually the best way of identifying windings, and factors of two for secondary voltages are significant as they may indicate centre-tapped sections, or 4 Ω and 16 Ω taps.

Uses and abuses of audio transformers

Transformers are among the most reliable of electronic components, often lasting 40 years or more, but they *can* be damaged. Transformers are made of wire that can fail if excessive current is passed, and insulation that can break down if it has to withstand too many volts.

The most common way of destroying an output transformer is to drive the amplifier well into overload so that one output valve switches off completely whilst the other is hard on. The leakage inductance of the half of the transformer associated with the switched off valve tries to maintain its current, and in doing so produces a very large primary voltage causing the interwinding insulation to break down:

$$E = -L_{\text{leakage}} \frac{di}{dt}$$

Since $di/dt \approx \infty$, the EMF developed is far higher than HT voltage, and it is easily capable of punching through the transformer interwinding insulation. If damp has been allowed to get at the transformer, then the (possibly paper) insulation will already be slightly conductive, and the possibility of breakdown is increased.

The author has still not managed to damage an output transformer, even when driving amplifiers to their full voltage output with only a very small

electrostatic loudspeaker connected across the anodes, but the possibility should be considered.

Guitar amplifiers and arcs

Since the rate of change of current at overload is high, and output transformers for guitar amplifiers are deliberately poor, implying a large leakage inductance, a sufficiently high voltage can be developed to strike an arc externally, even though the transformer may be designed to survive the experience. The voltage needed to strike an arc depends partly on the cleanliness of the path, so a dirty (conductive) path lowers the voltage, and a carbonized trail from a previous arc certainly reduces the voltage needed.

Although a high voltage is needed to strike an arc, once struck, it can be maintained by quite a low voltage. As an example, the xenon lamp used in a small cinema projector must be struck by a capacitive discharge of thousands of volts, yet may be maintained by only 26 V at 75 A. If an amplifier strikes an arc from the anode, it can only maintain the arc to a place that has a low resistance to ground, because a high resistance, such as a grid leak or cathode resistor, would limit the current, thus extinguishing the arc. The heater pins are connected directly to ground via the LT centre tap, so the most likely place for an external arc to strike is between anode and heater pins, because the only limiting resistance is the HT supply.

If we know that the amplifier will be thrashed, then a possible solution (depending on the amplifier) is to insert a resistor between LT CT and 0 V HT, perhaps a 4.7 kΩ 6 W W/W, in order to extinguish the arc. However, floating the LT supply may now cause hum problems because of poor heater wiring (routeing, dressing, connection to chassis).

Other modes of destruction

Excessive current through an output valve may cause thermal runaway from grid emission, melting the internal valve structure, thus dragging sufficient current through the output transformer that the primary winding fails. The simple cure is to keep your valve amplifier on display, and if a valve anode glows cherry red, switch it off immediately. (Output stages in valve amplifiers very rarely have fuses partly because the non-linear resistance of the fuse can cause distortion, but mostly because a fuse would not blow sufficiently quickly to protect the output valves.)

Small-signal transformers are usually damaged mechanically. They are fragile, and have windings of very fine wire that is easily broken. Treat them with respect.

Transformers screened in μ-metal cans must be handled carefully and not be dropped as the impact work hardens the μ-metal screen, greatly reducing its effectiveness (BBC 1:1 transformers designed to operate at −45 dBu had dire warnings on their screening cans about mechanical shock).

Magnetic core materials can deteriorate with time (this was a stock fault for the mains transformer of a particular picture monitor), and the author has recently seen a number of chokes and transformers whose aberrant behaviour can only be explained by poor core material. You might want to bear this in mind when choosing between an NOS part and a slightly more expensive new one.

Thermionic valves

History

The thermionic valve was not so much invented as discovered, and was a consequence of Thomas Edison's research into the extending the longevity of incandescent filament lightbulbs. It had been observed that as the lightbulb neared the end of its useful life, the glass became discoloured and darkened. (This effect is not often clearly visible on domestic lightbulbs, but it can be seen on non-quartz halogen torch bulbs and stage lamps.) The cause of the darkening was evaporation of the tungsten *filament* followed by deposition on the inner surface of the glass. In an attempt to counter tungsten evaporation, a *plate* was introduced into the (evacuated) *envelope*, and it was then noticed that if the plate was positively charged with respect to the filament, a current flowed across the vacuum. (Lightbulbs need a vacuum because the incandescent tungsten filament would otherwise oxidize so rapidly that it would burn.) Glass darkening can clearly be seen in the author's Ediswan R-type valve. See Fig. 4.24.

In 1904, John Ambrose Fleming[6] went rather further, and invented a new device that he termed an electrical *valve*, because it only allowed current to flow in one direction. His device used two carbon filaments, one of which was heated 'to bright incandescence of greater intrinsic brilliancy than if used as an incandescent lamp', although this was qualified later in the patent as only being ≈2000K (contemporary tungsten lightbulbs typically operate at ≈2900K). The other carbon filament, or *electrode*, was cold, and once a source of AC was connected between these two electrodes, he found that current could only flow in one direction. The new device was the *thermionic* diode, because of the *therm*al energy required to produce the *ion* flow.

Fig. 4.24 *R-type valve, note darkening of envelope. (The etched legend declares this valve to be 'Type approved by Postmaster General BBC')*

Strictly, only soft vacuum (or low pressure gas-filled) valves rely on the flow of gas ions – hard vacuum valves rely on the flow of electrons.

Although the diode had great curiosity value, and Fleming had suggested that it could be used for the detection of Hertzian (radio) waves, it was of limited practical use for two reasons. First, carbon emits electrons reluctantly, and second, electron emission is strongly dependent on temperature, which had to be kept quite low to avoid premature failure of the carbon filament. These two factors meant that the carbon filament had only ≈0.003% of the emission of a tungsten filament at 2900K, necessitating a sensitive mirror galvanometer to detect emission.

Lee de Forest's Audion patent[7] of 1908 interposed a platinum wire in the shape of a grid iron between the heated filament and the plate, and showed experimentally that amplification could be achieved. Although the patent reveals that he did not understand how it worked at the time, the new device was useful and quickly led to the commercial birth of radio. The amplifying

characteristics of the soft vacuum Audions were very variable, but the far more predictable hard vacuum triode soon evolved.

Emission

All metals have free electrons within their crystal structure, so some of them must be at the surface of the metal, but they are bound there by the nuclear forces between them and the adjacent atoms. However, the atoms and electrons are constantly vibrating due to thermal energy, and if the metal is heated sufficiently, some electrons may gain sufficient kinetic energy to overcome the attractive forces of the atoms and escape.

The heated metal in the valve is the cathode, and when this is heated to a temperature determined by the work function of the metal, an electron cloud or space charge forms above the surface of the cathode. Because like charges repel, the cloud eventually accumulates sufficient charge to prevent other electrons escaping from the surface, and an equilibrium is reached.

If we connect the plate, or anode, to the positive terminal of a battery, electrons are attracted from the cloud, and are accelerated through the vacuum to be captured by the anode. Because the electron cloud has been depleted, it no longer repels electrons as strongly, so more electrons escape the surface of the cathode to replenish the electron cloud.

Current cannot flow in the opposite direction because only the cathode can emit electrons, and only the positive anode can attract electrons.

Electron velocity

We mentioned that electrons were accelerated towards the anode, and this is quite literally true. At the exact instant that an electron leaves the cloud, it has theoretically zero velocity, but it is constantly accelerated by the electric field of the anode, and acquires energy proportional to the accelerating voltage:

$$E = eV = \frac{1}{2} m_e v^2$$

where: E = energy
e = electronic charge $\approx 1.602 \times 10^{-19}$ C
V = accelerating voltage
m_e = mass of electron $\approx 9.11 \times 10^{-31}$ kg
v = electron velocity

Rearranging, and solving for electron velocity:

$$\text{electron velocity} = \sqrt{2V \cdot \frac{e}{m_e}}$$

The ratio e/m_e is known as the electron charge/mass ratio, and has an approximate value of 1.7588×10^{11} C/kg. If we apply 100 V between the anode and cathode, the electrons will collide with the anode with a velocity of $\approx 6 \times 10^6$ m/s, or 13 million miles per hour.

Using the previous equation, it would appear that 512 kV (a common national distribution voltage) would be sufficient to accelerate the electrons to faster than light speed – which is an impossibility. The flaw is that the simple equation only accounts for the rest mass of the electron, but at relativistic (approaching the speed of light) velocities, the mass of the electron increases, thus requiring an infinite voltage to accelerate it to light speed. To account for this, the elegant equation given by Alley and Atwood[8] may be used:

$$\text{velocity} = c \cdot \sqrt{1 - \frac{1}{\left(1 + \frac{e}{m_e} \cdot \frac{V}{c^2}\right)^2}}$$

where c = velocity of light in a vacuum $\approx 2.998 \times 10^8$ m/s.

As an example of relativity at home, the display tube in a good quality colour television typically has a final anode voltage ≈ 25 kV, and therefore an electron collision velocity at the tube face of 202 million miles per hour, but the simple equation predicts a velocity 3.5% high.

Hospital X-ray machines deliberately fire high velocity electrons at a target, since X-rays are easily produced once the collision velocity exceeds 200 million miles per hour. Domestic display tubes do not therefore use a final anode voltage >25 kV, even though it would allow better focus, and sharper pictures.

Note that the distance between the anode and cathode does not feature in either equation because an infinite distance would also allow an infinite time for acceleration, and even if the rate of acceleration was very low, the collision velocity would still be reached.

Many effects within valves can be understood by having an appreciation of the collision velocity of the electrons as they hit the anode.

Individual elements of the valve structure

The cathode

Early valves betrayed their lightbulb origins, and were *directly heated* using a tungsten filament that was also the cathode. Tungsten is used in lightbulbs because it has the highest melting point of all electrical conductors (3695K, or 3422°C) and can therefore withstand the ≈3000K temperature necessary to generate white light. Although producing bright light was not actually necessary for early valves, it was quickly found that reduced cathode temperature caused electron emission to fall drastically, so early valves became known as *bright emitters*. The emitted current per unit area is:

$$I \propto T^2 \exp^{-\frac{e\phi}{kT}}$$

where: T = absolute temperature of the cathode = °C + 273.16
e = electronic charge ≈ 1.602×10^{-19} C
ϕ = work function of the cathode surface (≈4.55 for tungsten)
k = Boltzmann's constant ≈ 1.381×10^{-23} J/K
exp = base of natural logarithms ≈ 2.718

(See Appendix for the full theoretical Richardson/Dushmann equation.)

The emission efficiency of the cathode is important because not only does the filament dissipation increase the power requirement of the equipment, but the heat must be lost without damaging any other components. We therefore want to maximize electron emission for a given filament power, and the history of the cathode is thus concerned with the developing chemistry of the cathode emissive surface.

The first improvement was to use a *thoriated* tungsten cathode which not only had improved emission, but could operate at between 1950K and 2000K rather than 3000K. This reduced temperature was significant because valves primarily lose heat by radiation, and by Stefan's law:

$$E = \sigma T^4$$

where: E = power per unit area
σ = Stefan's constant ≈ 5.67×10^{-8} W/K^4/m^2
T = absolute temperature = °C + 273.16

Thus, 1975K only requires one fifth of the heater power to overcome the losses due to radiation compared to 3000K, and these valves are sometimes known as *dull emitters*. Although the emission had only been doubled, the reduction in heater power by a factor of five meant that the total improvement in emission efficiency was a factor of ten.

The real improvement came with the oxide-coated cathode, which operated at only ≈1100K, and was 100 times as efficient as the pure tungsten cathode. As an example, even in 1929, the P215 battery valve had a directly heated barium azide coated cathode requiring only 2 V at 150 mA. See Fig. 4.25.

Valve manufacturers often had proprietary cathode coating formulations, so the J.T.Baker Company manufactured the splendidly named 'Radio Mixture No. 3', which was composed of 57.3% barium carbonate, 42.2% strontium carbonate, and 0.5% calcium carbonate.[9] Sadly, this concoction almost certainly bears no relation to the apocryphal 'Love Potion No. 9'.

Unless the cathode is pure tungsten, the active emissive surface is only one molecule thick, and consequently fragile. The vacuum in a valve is never perfect, and there will always be stray gas molecules between the anode and cathode. A cold cathode prevents anode current, so zero voltage is dropped across the anode load resistor, causing V_a to rise to the full HT voltage. As

Fig. 4.25 *P215 directly heated oxide-coated cathode valve*

the cathode warms from cold, electrons are attracted towards the anode, but some collide with stray gas molecules to produce ions without significantly lowering V_a, thus repelling these ions towards the cathode with sufficient velocity to strip its surface. If this process of *cathode stripping* occurs sufficiently often, the cathode emissive coating can be significantly impaired, and oxide-coated cathodes are even more vulnerable than thoriated tungsten cathodes, so if $V_{a(pk)} > 2\,kV$ is required, thoriated tungsten cathodes are the norm. Because pure tungsten cathodes do not rely on a monomolecular layer for emission, they are almost immune to ion bombardment.

The cure for ion bombardment is to ensure that the vacuum is as nearly perfect as possible, and we will see how this is achieved when we look at the vacuum.

Another problem with oxide cathodes is cathode *poisoning*. If the cathode is kept at full operating temperature, but little or no current is drawn, a high resistance layer of barium orthosilicate forms at the interface between the barium oxide emissive surface and the nickel cathode structure. The interface resistance eventually reduces emission, but more significantly, it increases the noise generated by the valve.

Poisoned cathodes can occasionally be gradually recovered by operating the valve at a high anode current. Another method, often used on the cathode ray display tube used in traditional televisions, is known as *rejuvenation*,[10] and this works by temporarily increasing heater volts to heat the cathode to a higher temperature, and simultaneously drawing a large anode current. It should be realized that rejuvenation carries a risk of evaporating some of the cathode emissive surface, and contaminating the (nearby) control grid.

To avoid both cathode stripping *and* cathode poisoning, we can leave the heater supply permanently energized at 63% of full voltage, and only apply full voltage at full switch-on. There is nothing magical about the 63% proportion, it is simply a ratio that was found to extend the life, and reduce warm-up time, of the oxide-coated Plumbicon tubes used in the final generation of colour television cameras using tube sensors.

Thoriated tungsten filament fragility

An unpleasant side effect of thoriated tungsten cathodes is that they are very brittle, and so valves such as the 211, 300B, 813 and 845 should be handled with extreme care, and not be subjected to mechanical shock. The directly heated thoriated tungsten filament is a fine wire often shaped into a 'W' which is lightly tensioned by springs. See Fig. 4.26.

Unfortunately, thermal shock also kills thoriated tungsten filament valves. A 1994 study of transmitter valve longevity[11] found that each off/on cycle

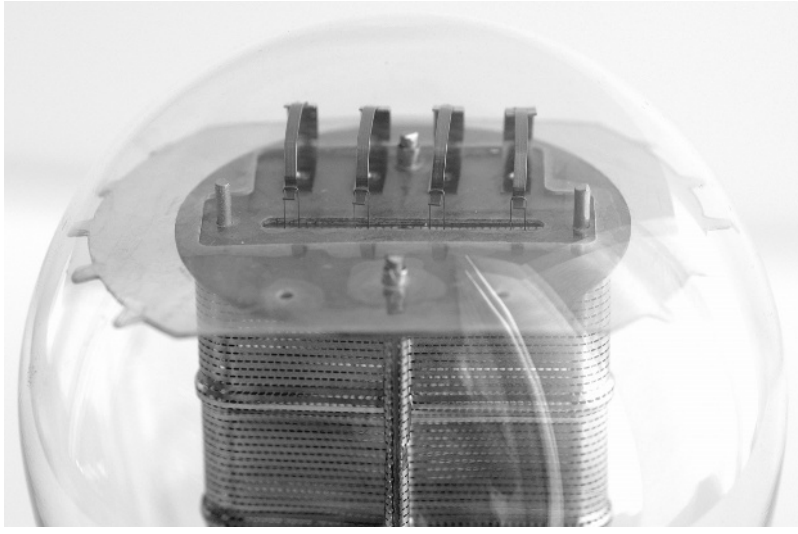


Fig. 4.26 *Tensioning a 'W' directly heated thoriated tungsten filament with leaf springs*

reduced filament life by 0.2% from its maximum life of 30 000 hours. This doesn't sound too bad, but it implies that 500 off/on cycles will destroy the filament, so if you switched the valve off and on every day, you could expect it to expire in less than 17 months. Unsurprisingly, the broadcasters took a dim view of this, and looked to see how life might be extended.

There are two reasons why the off/on cycle kills thoriated tungsten filaments:

- As the filament temperature passes through ≈900K, the Miller–Larson effect causes the grains of the metal to reorient themselves, so that the wire becomes thinner and longer. Worse, if a given section of the filament is slightly thinner, the increased current density causes increased localized heating which exacerbates the Miller–Larson effect and causes further necking of the filament. Eventually, this necking leads to such deep cracks that the remaining conductive material has sufficiently high current density and local heating to vaporize it, thus destroying the filament.
- The resistance of a cold filament is one tenth that of a hot one, so the cold inrush current is ten times higher than the operating current. The inrush current through the filament interacts with the earth's magnetic field to produce a small kick. Combined with Miller–Larson effect, this gradually deepens the surface cracks in the brittle filament. The damage done to the filament is proportional to the cube of inrush current, so a 'soft start' can be worthwhile.

If you had bought a quartet of NOS 845 at considerable expense, you would have a vested interest in avoiding Miller–Larson effect, and might want to permanently operate the filaments in standby mode at 80% of full voltage, and only apply full voltage at full switch-on, but note that this expends the emissive life at a rate of 1% of full filament voltage.

Directly vs indirectly heated cathodes

Early valves used lead–acid cells to power their directly heated cathodes, hence heater voltages that were multiples of 2 V. Having to take a heavy lead–acid battery to the local radio shop for periodic recharging was a nuisance, so later valves used the household AC mains supply. Unfortunately, AC heating caused audible hum owing to the following three mechanisms ranked in order of significance.

The thermal problem

The filament has to be made of sufficiently fine wire to give a high resistance that can be heated by a low current and not require excessively thick wires from the transformer. Because the wire is so fine, the thermal mass of the filament is low, and the temperature of the filament is partly able to track the applied power. This effect can be observed by noting that the thick filament of a car headlight dims slowly when switched off, yet the fine filament in a domestic AC mains lightbulb of the same power dims quickly. Because emission is modulated by heating *power* (V^2 or I^2), the mechanism produces hum at twice the AC mains frequency.

The electrostatic problem

The voltage drop across the filament affects V_{gk} because if we consider any one point to be connected to HT 0 V, other points are at different (and changing) potentials. If we connect the centre of the filament to HT 0 V, then the ends will be at equal and opposite voltages, so although one end of the filament emits more electrons, the other emits less. However:

$$I_a \propto V_{gk}^{\frac{3}{2}}$$

The significance of the three halves power law is that the excess of electrons drawn from the more positive end of the filament is *not* exactly nulled by the deficiency from the opposite end. The imbalance is most apparent when the

(AC) filament supply is at a positive or negative peak, so this mechanism also produces hum at twice AC mains frequency.

The thermal and electrostatic problems arise because the filament supply is a sine wave. A square wave filament supply would eliminate these problems, although preventing breakthrough of the higher harmonics into the audio circuitry via C_{hk} would be a major headache.

The electromagnetic problem

The magnetic field created by the filament heating current curves the flight of the electrons so that some miss the anode. When AC is applied to the filament, the direction of curvature alternates with the polarity of the filament current, so this mechanism produces hum at mains frequency.

Although various alternatives were tried, the best solution to these problems was the *indirectly heated* cathode,[12] whereby the emissive material was applied to a coated metal sleeve surrounding the filament, which was then called the *heater*. Although not explicitly stated in the patent, the key idea was that if the sleeve cathode had sufficient thermal mass, it would be unable to track the changing temperature of an arbitrarily thin heater wire. Because the metal sleeve did not pass heater current, its entire surface was at the same potential, so it was named a *unipotential* cathode, and this solved the electrostatic problem. If the sleeve is made from a magnetic material such as nickel, this tends to screen the filament's magnetic field, reducing the electromagnetic problem.

Since the sole purpose of the indirectly heated cathode is to reduce hum, the AC powering the heater must be electrically insulated from the signals on the cathode. Unfortunately, good electrical insulators also tend to be good thermal insulators, and the thermal resistance of the aluminium oxide electrical insulator means that the heater must be at 1650K to raise the cathode to 1100K, so indirectly heated cathodes need more heater power than directly heated cathodes. They also take longer to reach operating temperature, but for small-signal valves, the reduction in hum is invaluable, so the slow warm-up and loss of efficiency can be tolerated.

The cathode emissive surface is sprayed onto the outside surface of the nickel sleeve, and the heater reverts to pure tungsten. Even at this reduced temperature, the pure tungsten heater is still capable of limited thermionic emission, and the cathode looks like an anode to the heater filament. If a current was allowed to flow from the heater to the cathode, then this would add to the intended cathode to anode current, and hum would result. Fortunately, this problem is easily cured by superimposing a small (+10V seems sufficient) DC voltage on the heater supply, which then turns off the

unintentional diode between the heater filament and the cathode. RCA, who designed and made valves, recommended between +15 V and +40 V between the heater and cathode in many of the circuits that appeared in their publications.[13]

Despite all these efforts to eliminate hum, the heater filament could still induce hum into the signal circuitry either by leakage currents, or because of the imperfect magnetic shielding of the nickel cathode sleeve. In a further effort to reduce hum, the heater filament of the EF86 was wound as a helix in order to cancel the magnetic field caused by the heater current.

The only way to eliminate heater-induced hum is to use a DC heater supply with no AC content whatsoever, and this implies a *stabilized* supply, which has other benefits. Because cathode emission is so strongly temperature dependent, it is essential that the heater voltage is correct, and Mullard quoted a maximum permissible heater voltage variation of ±5%, which is exceeded by the current UK legal limit for mains voltage variation (+10%, −6%). A stabilized heater supply stabilizes the characteristics of the valve, and the elimination of thermal cycling of the cathode surface reduces low frequency noise.

As an aside, when the author installed an Automatic Voltage Regulator (AVR) to the test bench supplying his AVO VCM163 valve tester to combat mains voltage fluctuations, the AVR worked hardest between 4 pm and 11 pm. Is it a coincidence that the sound of the Hi-Fi seems to improve after midnight?

All of the final generation of small-signal valves use indirectly heated cathodes, and directly heated pure tungsten cathodes are now only used for high power transmitter valves.

Heater/cathode insulation

An indirectly heated cathode consists of a heater filament insulated by aluminium oxide folded and slid into a tightly enclosing tubular cathode. See Fig. 4.27.

No insulator is perfect, and they all deteriorate rapidly as temperature rises, which is unfortunate, since this particular insulator is red hot. Typically, the resistivity of aluminium oxide at its operating temperature is less than one millionth of its room temperature value. All heater/cathode insulation must therefore be electrically leaky, allowing leakage currents to flow between heater and cathode. Worse, if the insulation is contaminated, this imperfection produces 1/*f* noise. Irritatingly, one of the worst offenders for poor heater/cathode insulation is the otherwise excellent 12B4-A, so this valve

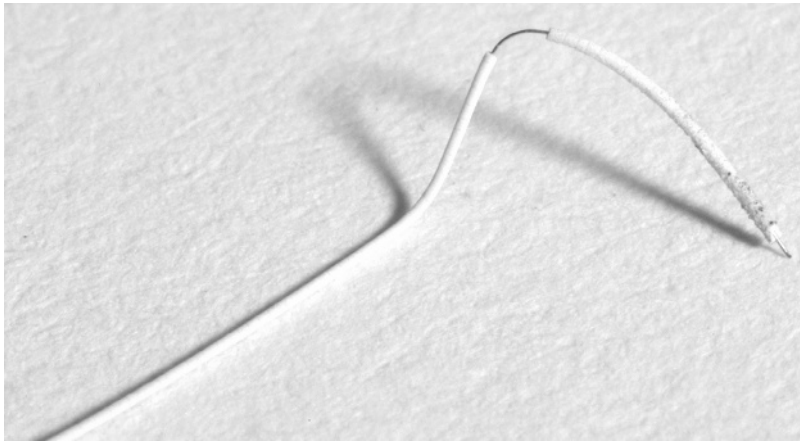


Fig. 4.27 *6K7 heater filament removed from cathode. The aluminium oxide insulation has been removed at the apex of the hoop to expose the much thinner tungsten filament*

must be screened to exclude those samples with poor (hot) insulation if noise is critical.

Heater power could be reduced, and the valve made more efficient, by reducing the thickness of the heater/cathode insulation, and this is exactly what was done in the transition from the International Octal-based generation to the later B9A generation, but this compromises electrical heater/cathode insulation.

Increasing the voltage across the heater/cathode insulation increases leakage currents. Although V_{k-h}(max.) is specified on data sheets as being anywhere from 90 V to 150 V (except for some ruggedized and 'P' series TV valves), this is a very 'soft' limit, since it is usually given at an arbitrary leakage current, nevertheless, a sufficiently high voltage *will* punch through the insulation to rupture the heater. Heater failure due to heater/cathode insulation breakdown is uncommon, but is most likely in cathode followers with high signal voltages or output stages with distributed loads (such as the McIntosh design).

Cathode temperature considerations

Because of the Richardson/Dushmann equation, electron emission, and therefore anode characteristics, are critically dependent on cathode temperature. Provided that anode dissipation is sufficiently low that it does

not further heat the cathode, cathode temperature is related to heater power (P) by:

$$T \propto \sqrt[4.5]{P}$$

Szepesi[14] also found that the oxide-coated cathode of the Tungsram HL4G produced minimum noise when operated at ≈1200K, and that a 60K drop in cathode temperature caused by a 25% drop in heater voltage doubled the noise power.

Operating oxide-coated cathodes at higher heater voltages dramatically shortens life because it increases evaporation of the emissive material, so $V_h \not> 105\%$. Thus, long life, low noise and stable anode characteristics demand heater supplies stabilized at the correct voltage.

Heaters and their supplies

It is usual to supply heaters in parallel from a constant voltage source (typically 6.3 V), or in series from a constant current source (typically 300 mA). If either type of supply drifts from its nominal value, undesirable changes in anode characteristics occur. Although 6.3 V regulators are easily made, linear regulators become increasingly inefficient as load current rises, and unless carefully designed and constructed, switched mode regulators can be electrically noisy. By comparison, a 300 mA constant current supply feeding a pure series heater chain is easily and efficiently implemented.

Valve manufacturers often specify series or parallel heaters, but is there actually any fundamental difference between the filaments of the two types, and could we use 6.3 V heaters (of equal current requirements) in a constant current chain?

Some 6.3 V valves were tested to see if there was any significant difference between the behaviour of their heaters. The valves were deliberately chosen to be as different as possible to magnify any difference between heaters. The 12AT7 was selected to be one whose heater flashed white at switch-on. See Table 4.1.

Within the limits of experimental error (which significantly worsened towards 1 V), the heater currents are in very close agreement, suggesting that parallel heater valves have essentially similar filaments. This is broadly to be expected, since tungsten is the only practical filament material.

A PL508 was then tested set to its correct current (300 mA), the voltage at that current measured, and a series of measurements was taken, which were normalized to 6.3 V and compared with the mean currents from Table 4.1. See Table 4.2.

Table 4.1 Percentage of normalized heater current against heater voltage for parallel (6.3 V) heaters

Heater voltage (V)	CV4024 12AT7 (0.30 A)	Mullard ECC83 (0.29 A)	Raytheon 5842 (0.30 A)	GE 6BX7 (1.45 A)	Mullard EL84 (0.79 A)	Mullard EL34 (1.475 A)	Mean	SD σ
6.30	100	100	100	100	100	100	100	0
6.00	97	97	100	97	96	97	97.3	1.37
5.50	92	93	93	93	92	93	92.7	0.52
5.00	87	90	90	88	87	87	88.2	1.47
4.50	83	83	83	83	82	85	83.2	0.98
4.00	77	78	80	78	77	78	78.0	1.10
3.50	70	66	73	72	72	73	71.0	2.68
3.00	67	60	67	66	67	67	65.7	2.80
2.50	58	62	60	61	60	60	60.2	1.33
2.00	52	55	53	54	51	53	53.0	1.41
1.50	42	45	47	45	43	44	44.3	1.75
1.00	30	34	37	37	32	34	34.0	2.76
0.50	20	21	20	26	19	20	21.0	2.53

A somewhat improved measurement technique was available when the PL508 was tested, nevertheless, the correlation between this and the parallel heaters is remarkable. There appears to be no significant difference between the filaments in valves specified for series or parallel heaters, and provided that individual heaters consume their correct power, there seems to be no reason why we should not mix the two types at will. This technique will be used in the μ-follower RIAA pre-amplifier described in Chapter 7.

Table 4.2 Percentage comparison of current heater valve (PL508) with mean current of voltage heater valves

Voltage	6.3	6.0	5.5	5.0	4.5	4.0	3.5	3.0	2.5	2.0	1.5	1.0	0.5
Av.	100	97	93	88	83	78	71	66	60	53	44	34	21
PL508	100	97	93	88	83	78	71	66	60	54	45	35	23

As a further test, the heater current of an EL34 was investigated at ½% intervals within a ±5% nominal voltage range, and the results plotted as a graph. An extremely close fit to a straight line was observed, indicating that the heater behaves as a constant resistance over this very limited range. Since $P = I^2R$ and $P = V^2/R$, two conclusions emerge.

First, parallel chains should be constant voltage (Thévenin) regulated, and series chains should be constant current (Norton) regulated.

Second, we should not mix topologies–series/parallel heater chains cause errors, because each heater no longer sees a perfect Thévenin or Norton source. As an example, because the 6SN7 double triode has its heaters internally wired in parallel it should *only* be used with a 6.3 V Thévenin supply, whereas a 12SN7 (whose heaters are in series) would ideally be powered by a 300 mA Norton supply.

As a more insidious example, a double triode initially tested on a valve tester with 6.3 V parallel heaters, and found to have perfectly matched anode characteristics between sections, would be mismatched by configuring the heaters in series unless the heaters were also perfectly matched. Matching should closely replicate the proposed conditions of use.

Heater voltage and current

Typical indirectly heated valves require approximately one minute to reach 99% heater temperature from cold, or 40 s when pre-heated at 80% current (63% voltage). When driven from a constant current source, heater terminal voltage is a very sensitive measure of heater (and by implication) cathode temperature. An International Servicemaster 14N7 ($I_h = 300$ mA) was tested by a 4½ digit DVM set to log heater voltage at 5 intervals. The results were normalized to 100% of the final heater voltage and are presented in Table 4.3.

As can be seen from Table 4.3, although the valve operates correctly within 60 s, a considerably longer time is required before the heater/cathode reaches thermal equilibrium. Since emission, and therefore valve operation, is temperature dependent, we cannot expect stable operation until five minutes after switch-on.

The most marked changes occur during the first minute of operation, so the heater voltage over the first 50 s of a selection of different types was tested using an HP54600B oscilloscope set to 20% V_h/div. vertically and 5 s/div. horizontally. See Fig. 4.28.

As can be seen, the valves all warm at different rates, hence the manufacturer's caveats about series heater chains. However, if the series heater

Table 4.3 Heater voltage against time when driven by a constant current

Time (seconds)	V_h (%)
0	22.62
5	36.28
10	46.60
15	58.93
20	78.75
25	89.29
30	93.14
35	95.24
40	96.87
45	98.06
50	98.80
55	99.19
60	99.38
90	99.54
120	99.70
150	99.80
180	99.87
210	99.93
240	99.98
270	100

chain is fed by a constant current source, unequal heating rates are not a problem.

Without changing oscilloscope settings, a Brimar ECC88 was pre-heated with 80% heater current, and the effect of this was compared with heating from cold. See Fig. 4.29.

Having established that valve heaters are essentially the same, is the series or parallel chain best in terms of heater power regulation? Irritatingly, parallel connection with a Thévenin source is rather better in terms of power regulation and thermal runaway, but a carefully controlled series connection with Norton regulation is usable, far more efficient (lower losses in the regulator), and has subsidiary advantages that will be investigated in Chapter 5.

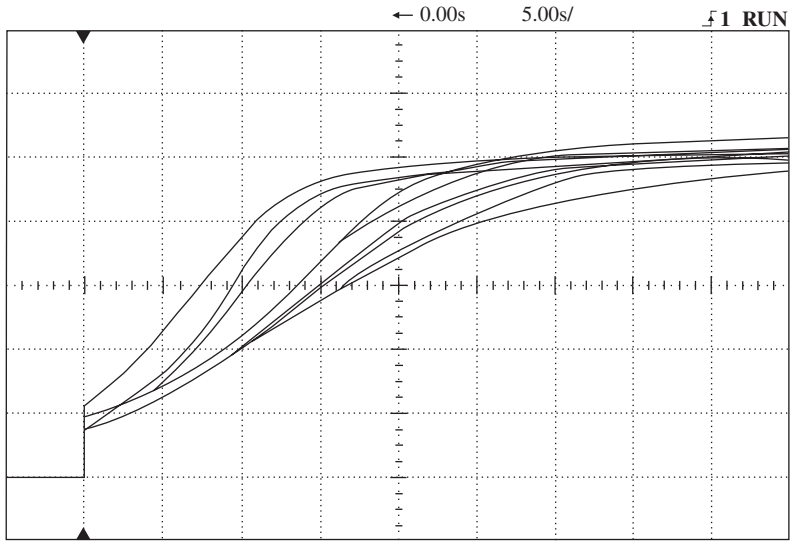


Fig. 4.28 *Heater voltage against time for various valves fed from constant current source*

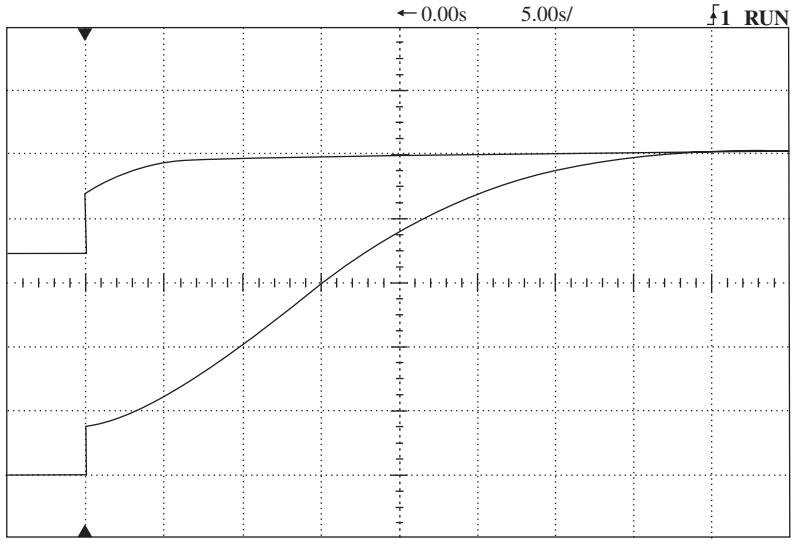


Fig. 4.29 *Heater voltage against time from constant current source. Upper trace: pre-heated. Lower trace: from cold*

The control grid

The control grid is wound from stiff fine wire (often tungsten) as a helix around the cathode, and is most effective close to the cathode surface, where the velocity of the electrons is low, than near the anode, by which time the electrons have acquired considerable momentum, and are not so easily repelled. Therefore, even in a valve having a succession of grids, the control grid is *always* the grid nearest to the cathode. The pitch of the grid winding and its positioning relative to the anode and cathode influence *gm* and μ. See Fig. 4.30.

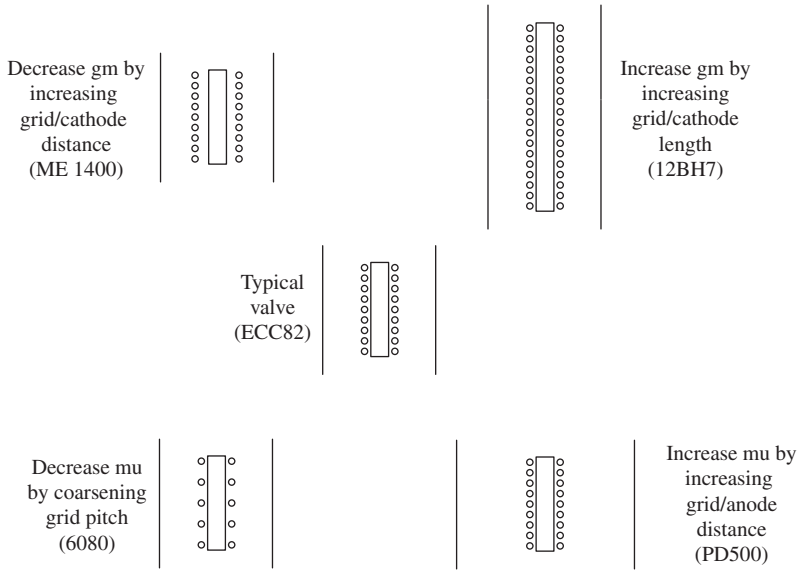

Decrease gm by increasing grid/cathode distance (ME 1400)
Increase gm by increasing grid/cathode length (12BH7)
Typical valve (ECC82)
Decrease mu by coarsening grid pitch (6080)
Increase mu by increasing grid/anode distance (PD500)

Fig. 4.30 *The effect of valve geometry on μ and gm*

As a practical example of the concepts shown in the diagram, two dissected valves were inspected under a travelling microscope. The 6080 ($\mu = 2$) has a grid pitch of $\approx$1.64 mm per turn, whereas the 417A ($\mu = 43$) has a pitch of $\approx$0.065 mm.

As an extreme example of the effect of anode–cathode spacing on μ, good quality valve televisions regulated their EHT supply because this avoided changes in picture size with brightness. Because their EHT was typically 15 kV, series regulators could not be used (heater/cathode insulation and efficiency problems), so shunt regulators were necessary. No valve has a perfect vacuum, so an increased anode–cathode spacing was needed to avoid arcing, resulting in humungous μ (1050 for the PD500).

Increasing *gm* requires that the grid be moved closer to the cathode, but if high μ is also required, then the grid winding pitch must be very fine, necessitating extremely fine wire for a uniform field. Thus, the WE416C ($gm \approx 65$ mA/V, $\mu \approx 250$) was specified to have a grid with a pitch of 1000 turns per inch (0.0254 mm), using 0.0003″ (12 μm) diameter wire, spaced 0.0005″ (20 μm) from the cathode.[15]

Because the control grid is so close to the cathode, a very small movement of the grid has a significant effect on the flow of electrons, and this is the cause of valve microphony.

Grid current

Although the control grid is normally a high resistance point, it can pass positive or negative grid current.

If we charge the grid positively with respect to the cathode, the grid reduces the repulsive effect of the space charge on electron emission at the surface of the cathode, and assists in pulling electrons away from the surface of the cathode. A much higher anode current flows, but some electrons are captured by the grid and flow out into the grid circuit, resulting in positive grid current, which drastically reduces input resistance. This is why Class AB2 output stages, which operate with positive grid current, are invariably preceded by a power driver.

Thermal runaway due to grid current

If the grid is allowed to emit electrons, negative grid current results, and depending on the value of the grid leak resistor and biasing, the potential of the grid may rise (*lowering* V_{gk}), causing an increase in anode current, and further heating the valve. The emissive material of the cathode then begins to evaporate, contaminating the grid and increasing grid emission. At worst, the grid may become so hot that it slumps and touches the cathode, completing the destruction of the valve, but increased valve noise is inevitable even if the valve is not actually destroyed.

Grid emission

Cathode stripping has been mentioned earlier as a problem for the cathode, but the stripped cathode material must go somewhere. The control grid is nearest to the cathode, so this sputtering process contaminates it with emissive material, greatly increasing the likelihood of grid emission.

The Richardson/Dushmann equation shows that there are two ways in which the emission of an uncontaminated grid may be minimized:

- Reduce grid temperature: Power valves cool their control grid by winding it on thick copper axial supporting wires that conduct the heat to radiant heatsinks at their ends. Note that a hot anode inevitably increases grid temperature.
- Increase grid work function: Electron emission is proportional to the inverse power of work function, so selecting a grid material with increased work function reduces grid emission.

Table 4.4 Comparison of emission for different grid materials

Metal	φ	Relative emission at 1100K
Tungsten	4.55	100%
Gold	5.28	0.045%
Platinum	5.63	0.0011%

As can be seen, a slight change in work function makes a huge difference to emission. Platinum gives by far the best performance, but a pure platinum grid would be expensive and too weak to be rigid. One practical solution was to platinum-clad molybdenum wire (25% platinum by weight[9]), but even this was expensive. Conversely, thinly gold plating a grid is cheap, so this technique is common (6080, 6S45П).

Frame-grid valves

If a horizontal brace is welded from one vertical control grid support rod to the other at both top and bottom, a frame is formed, so these are known as frame-grid valves. The advantage of this construction is that the grid wire can now be *tensioned* across the frame, greatly reducing sag, which enables far more precise grid cathode positioning of each wire, allowing closer spacing to the cathode, which increases *gm*. As example, the frame-grid E88CC comfortably achieves $gm = 10$ mA/V, whereas the traditional construction of the otherwise comparable ECC82 struggles to better 2 mA/V.

Variable μ grids and distortion

Variable μ valves are also known as remote cut-off valves, which refers to the gentle curve of their mutual characteristics, requiring an unusually large negative grid voltage to reduce I_a to 0. Radio receivers have to cope with a very large dynamic range of RF signals because they may be tuned from a strong local signal to a weak distant signal. To allow sufficient gain for the weak signal, but avoid overload on the strong signal, the RF gain of the receiver is made variable by an Automatic Gain Control (AGC) system. Typically, the AGC affects the gain of two or three valves simultaneously, and because the total gain is the product of individual gains, each valve only needs to change its gain by a small amount for the total gain to change significantly. The control grid is deliberately wound with an uneven pitch (which causes μ and *gm* to change with V_a), so these valves are known as variable μ triodes or pentodes.

Although variable μ valves are *designed* to change their gain with V_a, this causes distortion, but this is not a problem at RF because the signals are small, and distortion is proportional to amplitude.

In an audio triode, distortion is dominated by the variation of r_a with I_a, but once the loadline is flattened ($R_L >> r_a$) to lower distortion, we rely on constant μ, so the evenness of the grid pitch becomes important. Mechanically, the grid wire is wound and swaged into guide slots cut into the grid support rods whose position was determined by a lead screw. See Fig. 4.31.

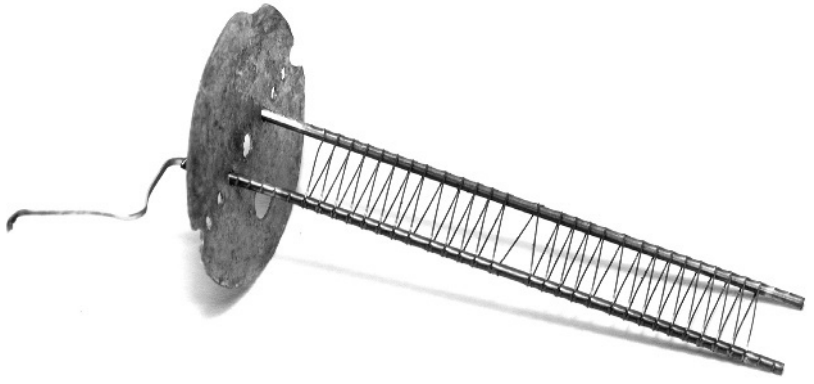


Fig. 4.31 *Control grid of 6K7 variable μ pentode. Note the deliberate deviations from constant pitch winding in the centre and a quarter of the way in from each end*

As the lead screw wears, backlash develops and the position of the guide slots becomes less precise, resulting in an uneven grid winding. Backlash is more of a problem with finer grid pitches (high μ valves), so it is more difficult to make a low distortion high μ valve than a low μ valve.

Reducing backlash and tightening production tolerances to reduce variation of μ is not an insurmountable problem – it just costs money, so this specification would be driven by the application. Fortuitously, it appears that the CV1988 (military 6SN7) required low variation of μ, so it has lower distortion than the commercial 6SN7.

Other grids

Tetrodes and pentodes have a helical screen grid (g_2) wound concentrically between the control grid and the anode, so a proportion of the cathode current is accelerated into the screen grid causing its temperature to rise. If the grid is coated with zirconium, this not only helps cooling by radiation, but it also absorbs residual gas, so the grid assists the getter in maintaining a good vacuum.

Beam tetrodes have beam forming plates rather than a helical wire grid. See Fig. 4.32.

Electrons are only weakly attracted to the suppressor grid (g_3) in a pentode, so this grid does not self-heat.

The anode

The anode is constantly bombarded by high velocity electrons. Although they have very little mass, their high velocity means that they possess considerable kinetic energy, which is converted into heat when they are stopped by the anode. An important specification for a valve is therefore the *anode* dissipation, because a hot anode heats the grid, causing grid emission. A secondary effect is that a hot anode releases gas, a phenomenon known as *outgassing*, which contaminates the vacuum.

Occasionally, some early valves used coarsely woven wire mesh anodes, which was claimed to 'prevent the grid becoming overheated by reflected radiation'.[16] See Fig. 4.33.

The mesh anode idea has been recently resurrected, although the 'mesh anode' 300B is actually pressed from metal sheet having rectangular holes stamped in it. See Fig. 4.34.

The anode dissipates heat by radiation, and in order to maximize radiating area, anodes often have fins. Another way to improve the loss of heat by

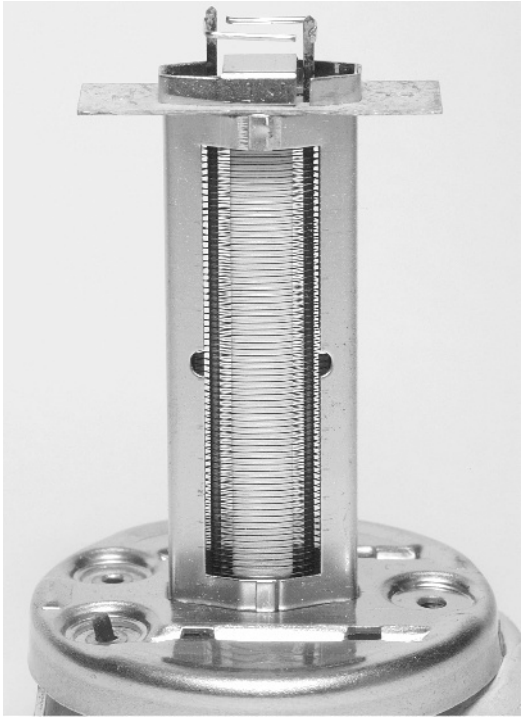


Fig. 4.32 *Beam forming plates and cathode/grid structure of QQV07-50 VHF dual beam tetrode*

radiation is to colour the anode black by coating the nickel anode with graphite. Even better, the 813 transmitting tetrode and 6528 series regulator dual triode use solid graphite anodes because they do not warp at high anode temperatures. The manufacturer of the 6528 made a virtue out of the very hot graphite anode by coating it with zirconium which has a great affinity for hydrogen, nitrogen and oxygen once heated above 800K (incipient red heat).

The anode's surroundings must be cool and capable of absorbing radiant heat. Otherwise, they will emit or reflect heat back to the anode, which is perfectly coloured to absorb radiant heat, thus raising anode temperature. Chrome-plated output transformers, etc. may look nice, but they could raise anode temperature. The worst possible surroundings for a valve would be a concentric chrome-plated cylinder, since this would focus the radiant heat back to the anode.

The EF86 low noise pentode has an electrostatic screen surrounding its anode to reduce hum, and on some examples, is formed from a shiny metal

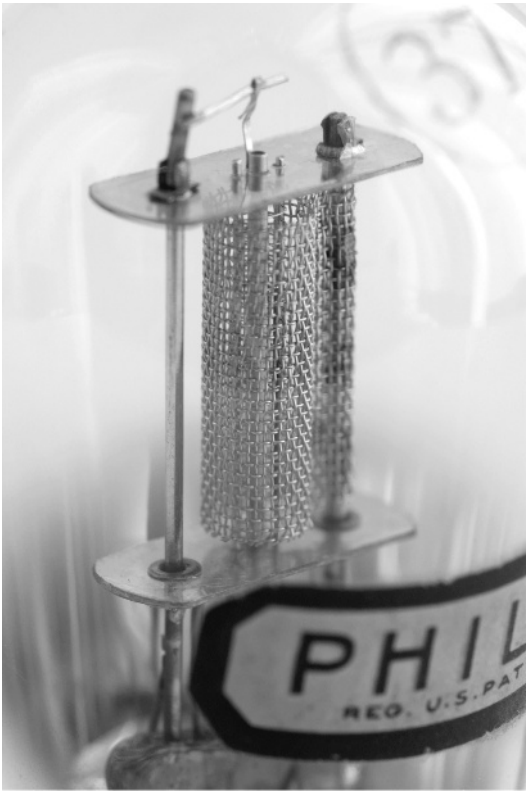


Fig. 4.33 *True mesh anode of a Philco 37*

sheet, but this should not be confused with the anode. This screen severely restricts anode cooling, but the *gm* of the EF86 is quite low anyway, so operating it at a high current (which would increase *gm*, but also increase P_a) would be pointless. The EF86 typically operates with a low P_a, so the electrostatic screen does not a cause a cooling problem.

Because electrons collide with the anode at high velocity, there is a possibility of dislodging more than one electron from the anode surface for each electron strike – an effect known as *secondary emission*. If the secondaries barely left the surface of the anode before returning, this would not be a problem, but if they stray any distance, they affect the electric field between the cathode and anode, causing distortion. The relative level of this emission is determined by the Secondary Emission Ratio (SER) of the material concerned. Nickel has a fairly low SER ($\approx$1.3), and together with its malleability, this is why it is commonly used for anodes and other pressed

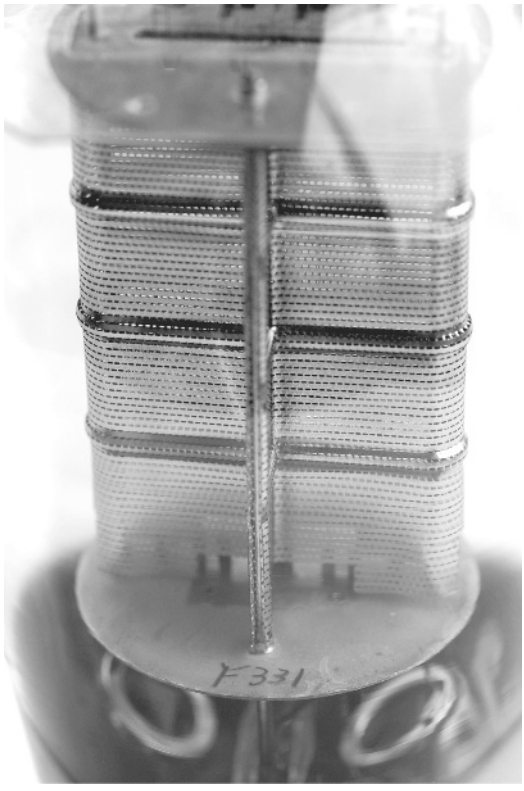


Fig. 4.34 *Punched plate 'mesh anode' 300B*

sheet valve electrodes. The SER of an electrode can be dramatically reduced by zirconium plating or by graphitizing the surface (coating it with colloidal graphite).

Graphite has a very low SER, but is rather fragile, so it can only be used in quite thick (>1 mm) structures. Moulded graphite anodes have increased thermal inertia, allowing $P_{a(peak)} >> P_{a(continuous)}$, so they are popular in transmitter valves such as the 813, 211 and 845. The Mullard QV08-100 has a massive anode structure formed out of two graphite slabs $\approx$8 mm thick, making an anode that is very tolerant of momentary overloads. See Fig. 4.35.

The vacuum and ionization noise

The quality of the vacuum within the valve is critical because initially uncharged gas molecules in the valve are likely to be struck by high velocity

Fig. 4.35 *Graphite slab anode of QV08-100 tetrode*

electrons on their way to the anode, possibly dislodging electrons to create positive gas ions. Positively charged ions are repelled from the anode, but are attracted to the grid/cathode structure, whereupon they are immediately discharged by a balancing number of electrons flowing up from the external paths to ground. Since the formation of ions and their subsequent discharge by the grid/cathode structure is random, the process creates random noise currents leading to *ionization noise*.

Ionization noise currents only become a problem when they flow through an external resistance such as a grid leak resistor. They then develop a voltage across that resistance in accordance with Ohm's law, and because valves are voltage operated devices (V_a or $I_a \propto V_{gk}$), the ionization noise voltage is amplified. If the grid resistance was zero, the ionization current would be unable to develop a noise voltage.

Low noise input stages use a high μ valve so subsequent stages do not degrade the noise performance, but a high μ valve implies fine grid pitch,

which greatly increases the probability of ions striking the grid rather than the cathode. Since R_g, the grid leak resistor, is invariably quite a high value, a significant noise voltage is developed and amplified. Because the grid effectively screens the cathode, very few ions strike the cathode, so cathode ionization current is greatly reduced, and because even an undecoupled R_k has a low resistance to ground compared with the grid leak resistor, this further reduces any noise voltage developed in the cathode circuit. Ionization noise in high μ stages is thus dominated by grid ionization noise current, and can be minimized by a low grid impedance to ground at audio frequencies, so transformer input coupling reduces the effects of ionization noise currents at low frequencies compared with capacitor coupling. (At very low frequencies, $Z_{sec.} \approx R_{DC(sec.)}$, which is quite low, whereas $X_C \approx \infty$, so capacitor coupling produces more $1/f$ noise.)

Output stages tend to use low μ valves to minimize Miller capacitance and preserve bandwidth, but low μ valves have a coarse grid pitch, biasing the probability of ion strike in favour of the cathode. Low μ valves operate with high bias voltages (V_{gk}), requiring high values of R_k. The combination of an increased proportion of ionization current and high R_k means that low μ valves should not leave their cathodes undecoupled if the effects of ionization noise currents are to be minimized.

A good vacuum is referred to as being *hard* whilst a poor vacuum is *soft*. Therefore, valves are sometimes described as having 'gone soft'. During manufacture, the air in the valve is pumped out, but some air will remain that cannot be removed by pumps, and the remaining gas is removed by the *getter*.

The getter

The getter is a metal structure often fitted near the top of the valve, coated with a highly volatile powder (usually a barium compound similar to the cathode emissive surface). Once the valve has been sealed and as much gas as possible has been pumped out, the getter is heated and the powder explodes, consuming the remaining gas. The force of the explosion throws molten barium onto the inside surface of the envelope to give the familiar mirrored coating at the top of the valve. The explosion is initiated electrically, either by directly passing a heating current through the getter's metallic supporting structure (metal envelope valves), or by shaping the getter as a short circuited turn and using transformer action to induce the heating current from an external RF source at ≈450 kHz (glass envelope valves).

Although some of the getter material is deactivated by the explosion, the getter must continue to consume gas molecules throughout the life of the valve

because gas continuously permeates the vacuum, either via leaks at the seals where the leads leave the envelope, or by outgassing from hot structures. The rate of a chemical reaction doubles with each 10 °C rise in temperature, so most valves thermally bond the getter to the hot anode with a thick wire. Because the anode is not at a constant temperature over its entire surface, and the wire has thermal resistance, some valves mount getters via short stubs onto the hottest parts of the anode. If bonded with only one stub, the mass of the getter combined with the spring of the stub can form a mechanically resonant system with a very high Q, so some special quality valves support their getters with two stubs to reduce the Q and hence microphony.

To be consumed by the getter, the gas molecules must touch it, and this is ensured by normal Brownian motion if the heater reaches operating temperature before HT is applied to the anode.

Soft valves can often be spotted by the gentle blue glow near the glass envelope which is due to the collision of ionized gas molecules with the glass. This should not be confused with the blue fluorescence that can be seen on the inner surface of the anode in valves such as the EL84, which is perfectly normal.

Valves stored for decades in a cold warehouse may have an imperfect vacuum because the getter was too cold to be fully effective. Fortunately, 24 hours in a domestic oven at 100 °C warms the getter and will often clear residual gas,[17] but beware that the phenolic bases of Octal valves can easily be damaged by a higher temperature. When first used, even new valves should have their heaters powered for a least half an hour before applying HT. Although a few hours of electrical use also clears the gas, the previous methods avoid the damaging ion bombardment suffered by the cathode until the getter has cleaned the vacuum.

The mica wafers and envelope temperature

The electrode structure, heatsinks and getter are supported and held rigidly in position by insulating mica wafers at the top and bottom of the anode structure. If this mica is not a perfect insulator, then leakage current paths will form, perhaps from the anode to the control grid, which would cause noise in a small-signal valve, but could cause destruction in a power valve.

When the getter is exploded, some molten metal may strike the mica wafer and make the mica slightly conductive. In order to lengthen the leakage current paths and increase their resistance, slots are cut in the wafer between the control grid and the anode.

Alternatively, the getter may be positioned such that it is less likely to spray onto the mica wafers, or the electrode supporting wafers may be screened by a sacrificial mica wafer or metal plate. The designers of the GEC KT88 not only used all of the above techniques to reduce leakage, but also made the electrode supporting wafers undersize, so that they did not touch the (assumed contaminated) envelope. See Fig. 4.36.

Fig. 4.36 *View of GEC KT88; note the measures taken to reduce leakage currents*

Even if the mica wafers have not been contaminated with conductive getter material, mica is not a perfect insulator, and like all insulators, its resistance falls with increasing temperature. The Sony C-800G studio condenser microphone uses a Peltier effect heat pump to cool the envelope of its pre-amplifier valve. Since the envelope is in contact with the mica wafers, they are also cooled, and it seems probable that the reduction in leakage currents

through the wafers, together with reduced anode temperature and consequent outgassing, is responsible for the reported improvement in noise.[18]

Not only does reduced envelope temperature reduce noise, it also improves valve life. To directly quote the Brimar Valves *Components Group Mobile Exhibition* (Nov. 1959) manual: 'The use of close-fitting screening cans of high thermal conductivity in intimate thermal contact with a large area of the bulb, in conjunction with an adequate heat sink can materially reduce the operating bulb temperature and very considerably improve the life of the valve.' Contrast this with a further comment from the same source: 'The use of screening cans which are not in thermal contact with the valve may seriously interfere with the cooling of the valve.'

A hot envelope implies a hot anode, and because the mica wafers support the anode, the anode conducts heat directly to the wafers. An excessively hot anode can heat the wafers to the point that they outgas water vapour (mica unavoidably contains water), which is particularly poisonous to oxide-coated cathodes.[19]

Micas can also poison cathodes because of vibration. A vibrating electrode chafes its support holes in the micas, producing fine mica dust.[20] Weight for weight, dust has a far greater surface area than a solid block, so it readily releases water vapour.

Ceramic support wafers are popular in transmitting valves such as the 845, and in ruggedized valves such as the 6384, because it avoids the water vapour problem.

Measuring envelope temperature

Many modern DVMs are capable of interpreting temperature from a Type 'K' (nickel chromium/nickel aluminium) thermocouple. Thermocouples for specific thermometer systems tend to be quite expensive, but a simple bead thermocouple on wires is cheap and perfectly usable. To measure envelope temperature, a suitable length of <0.7 mm bare copper wire can be looped around the valve and its ends *lightly* twisted to pull the bead reasonably firmly to the envelope (remember that the envelope is glass). See Fig. 4.37.

Note that envelope temperature is not constant over its surface and that the highest temperatures tend to be facing the centre of a flat anode or where a box anode is nearest the glass.

Valve cooling

The cross-sectional area seen from any point determines the heat radiated in that direction, so cylindrical anodes radiate their heat uniformly around the

Fig. 4.37 *Thermocouple bead clamped to valve by thin copper wire*

valve. However, thin, flat, box anodes can be almost invisible from one direction, making them more directional, so this effect can be exploited to steer heat away from sensitive areas.

Valve envelopes can be cooled by a push-on finned heatsink, but the additional cost is significant, so only critical equipment such as oscilloscopes or marine transmitters tended to use them. A clean, naked envelope loses much of its heat by radiation, but a finned heatsink in close thermal contact with the envelope seeks to convert all the heat losses to convection. Even in still air, this is a more efficient method of cooling (although it might only drop envelope temperature by ≈10 °C), but forced air cooling from a fan dramatically improves matters.

Push-on finned heatsinks are now readily available and should ideally be wiped internally with clear silicone, or molybdenum, grease to improve their

thermal conductivity to the envelope. *White* silicone grease would reflect radiant heat back to the anode!

Valve sockets – losses and noise

The BBC Equipment Department classified valve sockets into two types. Type 1 sockets used PTFE (PolyTetra-FluoroEthylene) or ceramic insulation, and Type 2 used phenolic. Type 2 sockets were not advised for use above 100 MHz, and equivalent resistance between adjacent pins at 30 MHz was stated to be not less than 750 kΩ and 100 kΩ for Type 1 and 2 respectively.

Out of curiosity, a few International Octal valve sockets from stock were tested for capacitance between adjacent pins. In balancing the bridge, not only was the capacitance determined, but also the '*d*' of that capacitance, which is inversely proportional to the leakage resistance, so a low value of '*d*' is desirable. See Table 4.5.

Although the ceramic sockets have slightly lower capacitance, the main difference is that phenolic sockets are typically ten times leakier, and will therefore increase noise currents from the anode into the grid. Although they

Table 4.5 Comparison of '*d*' for different valve bases

Type		C (pF)	*d*
Type 1: White ceramic	Unglazed PCB	1.3	≈0.01
		1.2	
		1.4	
		1.3	
	Glazed chassis	1.4	
		1.4	
		1.4	
		1.3	
Type 2: Phenolic	Black	1.7	0.08
		1.6	0.09
		1.6	0.095
	Brown	1.8	>0.1
		1.6	0.06
		1.7	>0.1
		1.6	>0.1
		1.9	>0.1

could not be compared directly in this particular test, PTFE sockets can be expected to have even lower leakage than ceramic, and are therefore thoroughly recommended.

The Loktal™ base

Despite its technical superiority over the established Octal base, the Loktal™ base introduced by Sylvania in 1938 was an unpopular diversion, and an evolutionary dead end, since the resulting valve was little smaller than an Octal valve. In order to produce a valve that could operate up to 225 MHz (7A4 versus the otherwise equivalent 6J5), capacitances, inductances and leakage paths had to be reduced. Eliminating the glass *pinch* within the valve and bringing electrode support wires directly to the pins shortened the valve, which reduced internal stray capacitances and lead inductance. Eliminating the phenolic base reduced leakage between pins, and the addition of an earthed metal base through which the pins protruded both screened and electrically guarded individual pins. The central metal spigot not only provided keying to ensure correct orientation, but had a ring that locked the valve into the base once inserted. In order to avoid trademark infringement, competing companies changed the nomenclature to 'Loctal', or referred to them as 'Lock-in' valves and designated the base B8G.

For high quality audio, Loctal valves are excellent, since their electrode construction (and consequent distortion) is the same as the Octal generation, but they have the obvious technical superiority of the Loctal base. Since Loctal valves were designed for use above 100 MHz (when socket losses become significant), PTFE insulated Loctal bases were made, and NOS bases are currently available.

The 1939 Sylvania manual misleadingly specifies Loktal™ heater voltage as being 7 V or 14 V for 130 V line, allowing them to use the new '7' and '14' prefixes for their type designation, but this is directly equivalent to 6.3 V or 12.6 V for 117 V – their nominal line voltage, so these valves actually have standard heaters. The contrived prefixes were simply a marketing ruse to distinguish the new base.

The glass envelope, and the pins

The envelope maintains the vacuum within the valve; careless handling will crack the valve, and air will enter. The easiest way to crack the envelope is to bend the pins whilst attempting to insert the valve into a new socket. It is therefore good practice to plug old valves into the new sockets of a new

amplifier, and test the amplifier before inserting the expensive new valves. This then spreads the fingers of the socket slightly, and the new valves can be inserted without fear of damage. A hidden problem is that repeated plugging and unplugging can create micro-fractures in the glass near the pins, allowing just sufficient gas to leak into the valve to subtly degrade the vacuum, and increase ionization noise (for this reason the British Standards Institution[21] warned against repeated testing of special quality valves). A terminally damaged envelope is easily spotted, because the mirror coating due to the getter turns white.

If the envelope is allowed to accumulate fluff, it will be thermally insulated, and the valve will run hotter, with all the dire consequences that entails. Valves should be clean and shiny to promote long life.

Not all electrons accelerated towards the anode are captured, and so a charge can build up on the inside surface of the (insulating) glass envelope. Coating the glass with graphite renders it conductive, and connecting it to the cathode allows the charge to be drained away.

The pins are made of Kovar, which is an iron–nickel–cobalt alloy having the same coefficient of thermal expansion as the glass envelope in order that leaks do not occur at the seals as the valve warms. If stored under damp conditions, Kovar can rust, so valves should ideally be stored in evacuated plastic bags. Loctal valves were the earliest valves to be introduced with Kovar pins, and their unpopularity at the time means that a 50 year old NOS valve is not uncommon, so beware of this problem. If you have a good stock of valves, store them in a warm dry place – not in a shed or garage.

On some valves, these pins may be gold plated, but this plating will be quickly removed by repeated plugging and unplugging. Gold plated pins used to be a sign of quality (although Brimar did not always bother to gold plate their excellent E88CC), but some modern valve manufacturers cheerfully gold plate selected valves sourced by their standard production line, whereas traditional Special Quality (Mullard) or Trustworthy (Brimar) valves were consciously designed/produced to be better, rather than selected from a standard production line.

Although gold is corrosion resistant and assists in achieving a good contact with the socket, silver has better conductivity ($\rho = 1.47 \times 10^{-8}\,\Omega$/m compared with $2.05 \times 10^{-8}\,\Omega$/m for gold and $1.54 \times 10^{-8}\,\Omega$/m for copper). As frequency rises, skin effect causes conduction to occur principally at the surface of the conductor, so the MUSA video connector originated at Post Office telecommunications exchange London *MU*Seum *A* was silver plated to improve conductivity. Similarly, transmitting valves intended for use at VHF sometimes have silver plated pins. Sadly, unattended silver corrodes quite badly.

PCB materials

Glass Reinforced Plastic (GRP) boards are rather less than ideal for valve audio because of leakage resistance. The leakage occurs because the epoxy resin does not always seal perfectly to the glass fibres and surface tension draws water vapour into the resulting gaps never to be released. Many years ago, the author built circuits on Synthetic Resin Bonded Paper (SRBP) and felt that they sounded better than the same circuit built on GRP, but at the time could not see any engineering reason why, and put it down to imagination. The crucial difference between GRP and SRBP is that SRBP is porous over its entire surface and not just at its edges. SRBP can therefore lose water vapour over its entire surface as it heats, whereas a GRP board can only lose water vapour at its edges. A warm SRBP board could therefore actually be less leaky (even though nominally a poorer material) than a warm GRP board. Because water is polar, the problem of dielectric leakage becomes even more acute at high frequencies, so microwave and >200 MHz oscilloscope designers have avoided the material for decades, preferring to use PTFE.

References

1 Scroggie, M. G. (1971) *Radio and Electronic Laboratory Handbook*, 8th ed. Iliffe, p. 475.
2 Hughes, L. E. C. and Holland, F. W. (eds). (1967) *Electronic Engineer's Reference Book*, 3rd ed. Iliffe.
3 Thiele, A. N. (1976) Air-cored Inductors for Audio. *Journal of the Audio Engineering Society*. June, Vol. 24, No. 5, p. 374.
4 McIntosh, F. H. and Gow, G. J. (1949) Description and Analysis of a New 50-watt Amplifier Circuit. *Audio Engineering*. December.
5 Sowter, G. A. V. (1987) Soft Magnetic Materials for Audio Transformers: History, Production, and Applications. *Journal of the Audio Engineering Society*, October, p. 760. Vol. 35. No. 10.
6 British Patent No. 24,850 (1904) Improvements in Instruments for Detecting and Measuring Alternating Electric Current.
7 US Patent No. 879,532 18th February 1908.
8 Alley, C. L. and Atwood, K. W. (1962) *Electronic Engineering*, Wiley.
9 Rosebury, Fred. (1965) *Handbook of Electron Tube and Vacuum Techniques*, Addison-Wesley.

10 Upton, Lane S. (1973) The Rejuvenation of Vacuum Tubes. Old Timer's Bulletin. December. Available at: http://www.antiquewireless.org/otb/rejuve.html
11 Johnson, Walter (VOA). (1994) Techniques to Extend the Service Life of High-Power Vacuum Tubes. NAB Engineering Conference Proceedings. National Association of Broadcasters. pp. 285–293.
12 British Patent No. 209,415 (1924) Improvements in Cathode Structures for Vacuum Thermionic Tubes.
13 RCA (1959) *Receiving Tube Manual, Technical Series RC-19*.
14 Szepesi, Z. (1939) The Temperature Response of the Shot Effect of Valves with Oxide-coated Cathode, *The Wireless Engineer*, February, pp. 67–70.
15 Western Electric. (1971) 416C Valve Data Sheet.
16 Henderson, F. E. (GEC) (1943) Introduction to Valves. 2nd ed. *Wireless World*.
17 Jones, Morgan. (2000) New Life for Old Valves. *Electronics World*, November, pp. 863–867.
18 Spencer-Allen, Keith. (1992) Chill Factor. *Studio Sound*, November, pp. 56–59.
19 Bendix Corporation. 6384 Valve Data Sheet.
20 Rowe, E. G. and Welch, Peter. (1954) Developments in Trustworthy-Valve Techniques. *Electrical Communication*, September, pp. 172–188. Available at: http://www.thevalvepage.com
21 British Standard Code of Practice CP1005. BSI (1962) The Use of Electronic Valves.

Further reading

Chaffee, E. Leon (1933) *Theory of Thermionic Vacuum Tubes*, McGraw-Hill.
Duncan, Ben (1996) *High Performance Audio Power Amplifiers*, Newnes.
Eastman, A. V. (1949) *Fundamentals of Vacuum Tubes*, 3rd ed. McGraw-Hill.
Kaye and Laby (1995) *Tables of Physical and Chemical Constants*, 16th ed. Longman.
Thrower, Keith R. (1992) *History of the British Radio Valve to 1940*, MMA International.
Tomer, Robert B. Sams (1960) Getting the Most Out of Vacuum Tubes. Reprinted by Audio Amateur Press (2000).

Vyse, Barry and Jessop, George (2000) *The Saga of Marconi Osram Valve*, Vyse Ltd.
Williams, Tim (1991) *The Circuit Designer's Companion*, Butterworth-Heinemann.

5

Power supplies

A power supply is a device that converts one voltage to another, more convenient voltage, whilst delivering power.

Valve amplifiers need a DC High Tension (HT) supply and one or more heater, or Low Tension (LT) supplies, which may be AC or DC. Often, the supplies for the pre-amplifier and power amplifier will be derived from the same power supply, which is frequently integral to the power amplifier, but this need not be so.

In this chapter, we will identify the major blocks of a power supply, see how to design them, then design a pair of complete supplies.

The major blocks

There are two fundamental types of power supply, *linear* and *switchers*. See Fig. 5.1.

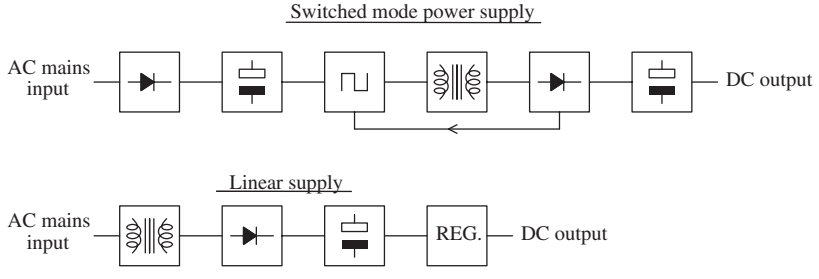


Fig. 5.1 *Comparison of linear and switcher power supplies*

In a switcher, the AC mains input is rectified, switched at a high frequency, typically >50 kHz, transformed, rectified, and smoothed; regulation is part of the switching element. Switchers are small, light and efficient. Their design is highly specialized, and early designs generated copious RF noise, but designs conforming to modern Electro-Magnetic Interference (EMI) standards are surprisingly quiet, and can be useful for heater supplies.

By contrast, a linear supply transforms the 50 Hz AC mains directly, requiring a bulky mains transformer. This is then rectified by valves or semiconductors, smoothed by large capacitors, and possibly even larger inductors, and then regulated if necessary. Linear supplies are heavy, inefficient, easily designed, but low noise. Valve amplifiers use lots of them, so we had better know how to design them.

Power supplies are designed from the output back to the input. Since they are designed after the amplification stages, it is tempting to think of them as an afterthought; indeed, some commercial products reflect this attitude. It is most important to realize that an amplifier is merely a modulator, and controls the flow of energy from the power supply to the load. If the power supply is poor, then the most beautifully designed amplifier will be junk.

Rectification and smoothing

Whilst we may not have a regulator on the output of the supply, we will always have rectification and smoothing. The two functions are inextricably bound together, and determine the specification of the mains transformer, so this is the starting point for design. Since we need to rectify the sinusoidal AC leaving the transformer with maximum efficiency, we will only consider full-wave rectification. Half-wave rectification is not only inefficient (because it only uses alternate half cycles), but it also causes DC to flow through the transformer, and even small DC currents can cause core saturation. A saturated core is lossy and produces leakage flux which can induce hum currents into nearby circuitry.

Choice of rectifiers/diodes

There are two forms of full-wave rectification; the centre-tapped rectifier, and the bridge rectifier. See Fig. 5.2.

The bridge rectifier is the usual modern topology because it economizes on transformer windings. The centre-tapped rectifier was traditional in valve circuits because it economized on rectifiers (which were expensive).

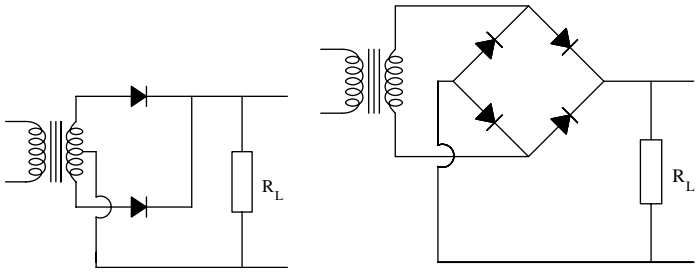


Fig. 5.2 *Full-wave rectification*

When we consider HT supplies producing $V_{DC} < 1\,kV$, we have a choice between silicon, or hard vacuum thermionic rectifiers such as the GZ34. Valve rectifiers are inefficient. Not only do they need a heater supply, but they also drop tens of HT volts across themselves and increase the supply's output resistance. They are fragile in terms of ripple current (which we will discover in a moment), and therefore there is a maximum value of capacitance that can be connected across their terminals. Even then, the total resistance in series with each anode must exceed a minimum value:

$$R_{series} \geq R_s + n^2 R_p$$

where: R_s = secondary resistance
R_p = primary resistance
n = secondary to primary turns ratio

Although the table allows a quick comparison between the common double diode rectifier valves, for detailed design we must refer to a manufacturer's data sheet.

	$I_{DC(max.)}$	R_{series} ($V_{out} = 300\,V$)	$C_{(max.)}$	I_{heater}
EZ90/6X4	70 mA	520 Ω	16 µF	600 mA
EZ80/6V4	90 mA	215 Ω	50 µF	600 mA
EZ81/6CA4	150 mA	190 Ω	50 µF	1 A
GZ34/5AR4	250 mA	75 Ω	60 µF	1.9 A
GZ37	250 mA	75 Ω	60 µF*	2.8 A

* Mullard did not specify $C_{(max.)}$ for the GZ37, but $i_{a(pk.)} = 750\,mA$ for both the GZ34 and GZ37, so $C_{(max.)} = 60\,\mu F$ may be assumed.

New Old Stock (NOS) Mullard GZ34 are now very scarce, and therefore expensive, whilst some modern GZ34 have been reported to be fragile at high voltages, so the current-hungry NOS GZ37 is a popular substitute. The indirectly heated EZ80 and EZ81 are cheaper and more readily available, and are ideal for pre-amplifiers or small mono power amplifiers. The unpopular EZ90 is not as well specified as the EZ80, but is likely to be even cheaper.

Indirectly heated rectifiers are designed to operate from the same 6.3 V heater supply as the signal valves, but this means that $V_{hk} \approx 300$ V, which puts a strain on the heater/cathode insulation, implying noise currents from the rectifier's cathode into the grounded common heater supply. If low noise is paramount, we could transfer the stress from the fragile heater/cathode insulation to the more robust mains transformer by providing a dedicated rectifier heater supply tied to the rectifier cathode.

Hard vacuum valve rectifiers have only one clear advantage over silicon, but this advantage may be sufficient to make us tolerate their foibles. The *rise time* (time taken to change from 10% to 90%) of their output voltage when fully loaded ≈5 s, which greatly reduces the inrush current of electrolytic capacitors in comparison with semiconductor rectifiers. See Fig. 5.3.

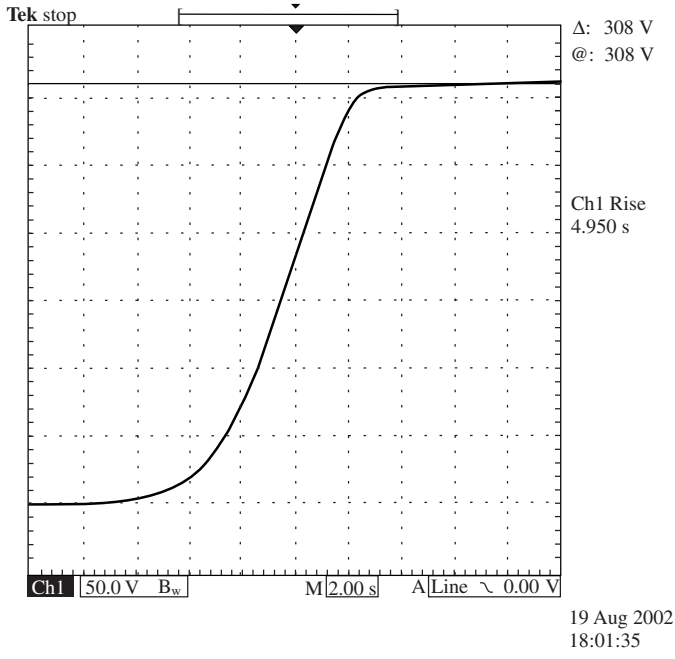


Fig. 5.3 *Gentle rise of HT supplied by EZ81 rectifier with 120 mA load*

Devotees of hard vacuum valve rectifiers point out that the valve switches on and off more cleanly than silicon, thus exciting fewer resonances in the power supply, but the author's experience is that both types of rectifier produce switching spikes and that it is the smoothing/snubbing arrangements that are important. If there is an improvement due to valve rectifiers it is more likely to be due to the reduced ripple current that they enforce (see later).

Whichever rectifier topology we choose, we must ensure that it is capable of withstanding the stresses imposed upon it by the surrounding circuit. When considering low frequency rectification derived from the mains, we need only specify the voltage and current ratings. However, neither of these ratings is quite as straightforward as it might seem. See Fig. 5.4.

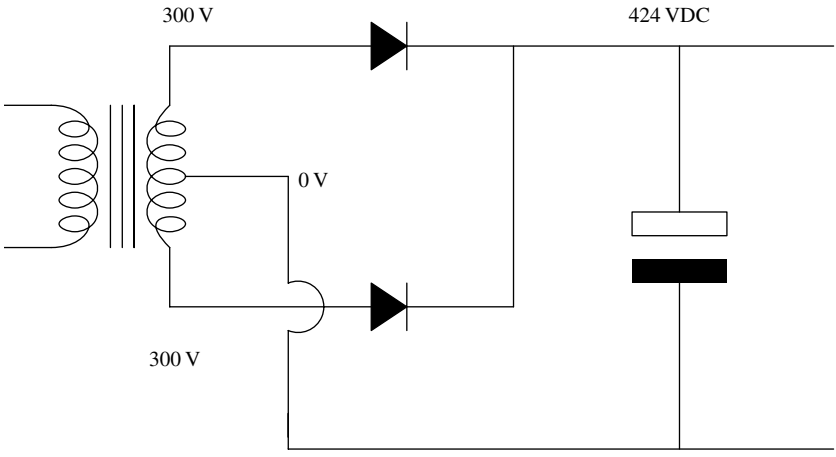


Fig. 5.4 *Effect of capacitor on rectifier ratings*

The diagram shows a centre-tapped rectifier using silicon diodes to rectify the 300–0–300 V secondary. The off-load voltage at the reservoir capacitor will be 424 V_{DC} (note that this is much higher than if a valve rectifier had been used – silicon diodes and valve rectifiers are *not* directly interchangeable). The diode voltage rating that concerns us is the reverse voltage rating, known as V_{RRM} (Reverse Repetitive Maximum), also known historically as PIV (Peak Inverse Volts).

The following table compares required diode voltage ratings for the two topologies.

	Diode rating: (V_{RRM}/V_{RMS})	V_f
Centre tapped	$2\sqrt{2}$	1
Bridge	$\sqrt{2}$	2

When rectifying high voltages, the centre-tapped rectifier has the disadvantage that it needs diodes having double the V_{RRM} rating. Thus, the 300–0–300 V transformer with centre-tapped rectifier requires diodes having $V_{RRM} > 849$ V, but a single 300 V winding with a bridge rectifier could produce the same output voltage, and only require diodes having $V_{RRM} > 424$ V. Nevertheless, the centre-tapped rectifier is very useful in low voltage/high current silicon circuits because of the lower forward diode drop, only 1 V_f, compared to 2 V_f for the bridge.

Valves such as GZ34, EZ81, EZ80, etc. that are intended for use in centre-tapped rectifiers normally require a centre-tapped transformer, but a hybrid valve/semiconductor rectifier circumvents this problem.[1] See Fig. 5.5.

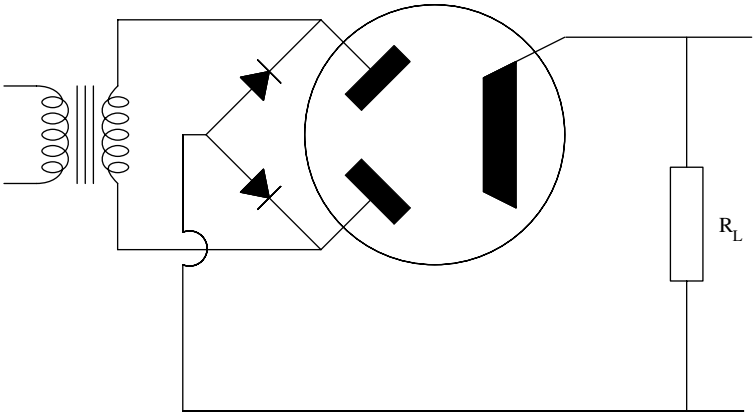


Fig. 5.5 *Hybrid valve/semiconductor rectifier*

When the diode feeds a reservoir capacitor, pulses of current many times greater than the DC load current flow. Fortunately, modern silicon diodes are designed with these peaks in mind, and for full-wave rectification it is usually sufficient to choose a current rating for each diode equal to half the DC load current. (This is because any given diode in a full-wave rectifier is only switched on for half the time.)

Mercury vapour rectifiers

Mercury vapour rectifiers have recently become fashionable, and their soft blue glow is certainly very attractive. Mercury vapour rectifiers are fragile and poisonous, so they must be treated carefully both electrically and mechanically. They rely on vapour, yet the mercury quickly condenses as metallic droplets onto the envelope, so the heater must first evaporate the droplets, which requires that they are mounted vertically. The cathode requires pre-heating before HT can be safely applied, as shown in the table.

Pre-heat time	**Ediswan**[2]	**Mullard**[3]
After storage/mechanical disturbance	>15 minutes	>30 minutes
Day to day	>60 s	>60 s

To avoid flashback, mercury vapour rectifiers are typically only rated to operate between 20 °C and 60 °C, although some can only tolerate ≤ 50 °C, so a fan could be needed to disperse hot air from other components. In addition, the 866 rectifier requires a bleeder resistor drawing 10% of the full load current across the output of the rectifier.

Mercury vapour rectifiers have a lower forward drop ($\approx$15 V) and much lower internal resistance than hard vacuum rectifiers, and are available in higher voltage and current ratings. However, they switch on and off very much more abruptly than their vacuum counterparts, so they are likely to oscillate if the anode lead is not snubbed with lossy ferrite beads or RF chokes, and may need to be enclosed by a metal screening can. The best way to detect oscillation is not with an oscilloscope, but by moving an AM radio nearby and listening for a buzz as the radio approaches the rectifier. Mercury vapour rectifiers combine the disadvantages of silicon and valves by needing heater supplies, HT delay circuitry and snubbers to give electrical performance very slightly better than silicon. (No charge storage to cause current overshoot.)

RF noise

Rectifiers are switches. Although the following argument implies a resistive load on the rectifier, the results are also valid for the load presented by a reservoir capacitor.

As the input AC waveform rises through 0 V, one or more diodes will switch on, and stay switched on until the waveform falls through 0 V, when the other diode, or diodes, switch on. All diodes need a minimum forward bias before they will switch on, even if it is only the 0.7 V required by silicon. This means that there is a dead zone symmetrically about 0 V when no diodes conduct. The transformer, which is inductive, has been switched off, and it tries to maintain the current flow, but in doing so, it generates a voltage:

$$V = -L\frac{dI}{dt}$$

Fortunately, there are usually plenty of stray capacitances within the transformer to prevent this voltage from rising very far, but if we are unlucky, the shock applied to the system may excite a resonance resulting in a damped train of oscillations. Using a search coil, the author once observed bursts of 200 kHz leaking from a mains transformer for this very reason. Happily, this problem can be cured simply by bypassing each individual diode with a 10 nF film capacitor having a voltage rating equal to the diode V_{RRM} rating.

Whether we use a bridge rectifier or a centre-tapped rectifier, we still apply the same waveform to the succeeding circuit. The waveform, although it is of only one polarity, is not a smooth DC. The function of the smoothing element is to reduce the ripple, either to a satisfactory level, or to a level such that a regulator can cope with it.

The single reservoir capacitor approach

The simplest way of smoothing the output of the rectifier is to connect a reservoir capacitor across it, and feed the load from this reservoir. See Fig. 5.6.

Assuming no load current, the capacitor charges to the full peak value of the AC leaving the transformer $(V_{sec} \times \sqrt{2})$.

Ripple voltage

The output of the rectifier tops up the charge in the capacitor every cycle, so that at the peak of the waveform, the capacitor is fully charged. The voltage from the transformer then falls away sharply, so the rectifier diodes switch off. Load current is now supplied purely from the capacitor, which discharges exponentially into the resistive load until the transformer output voltage rises sufficiently to recharge the capacitor, and restart the cycle. See Fig. 5.7.

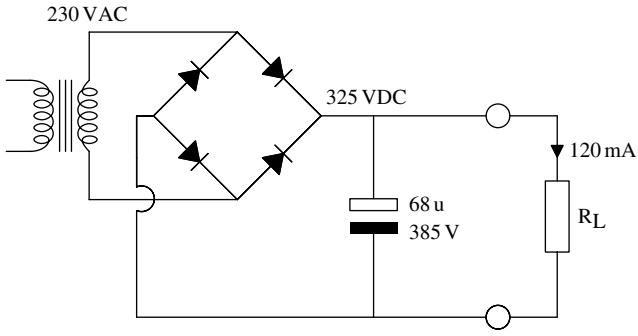


Fig. 5.6 *Power supply using reservoir capacitor*

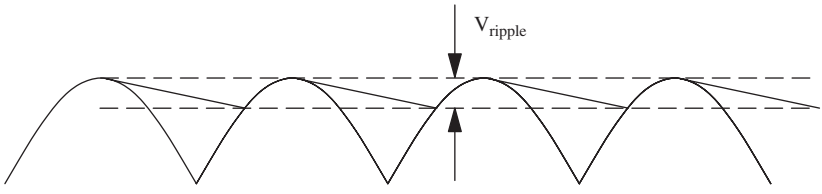


Fig. 5.7 *Ripple voltage across reservoir capacitor caused by charge/discharge cycle*

Although the reservoir capacitor theoretically discharges exponentially, for any practical value, the discharge curve may be taken to be a straight line. (If the load is a series regulator, the discharge curve truly *is* a straight line.) Given this approximation, it is easy to calculate what the output ripple voltage will be.

The charge stored in a capacitor is:

$$Q = CV$$

The total charge, due to a current I flowing for time t, is:

$$Q = It$$

We can combine these equations:

$$CV = It$$

Rearranging:

$$V = \frac{It}{C}$$

This equation gives the voltage *change* on the capacitor due to the capacitor supplying current I for time t. If mains frequency is 50 Hz, then each half cycle is 0.01 s. If we now make another approximation, and say that the capacitor supplies current *all* of the time, then $t = 0.01$. We now have a useful equation:

$$v_{\text{ripple(pk-pk)}} = \frac{0.01I}{C}$$

It might be thought that this equation is of little use, since two sweeping approximations were used to derive it, but as the reservoir capacitor is nearly always an electrolytic capacitor, whose tolerance could be ±20%, we need a very inaccurate equation before it compares to the error introduced by component tolerances!

We can now calculate the ripple voltage at the output of our example circuit in Fig. 5.6 which had a 68 μF capacitor, and a load current of 120 mA.

$$v_{\text{ripple}} = \frac{0.01 \times 0.12}{68 \times 10^{-6}} = 18\,V_{\text{pk-pk}}$$

which is about 5% of full voltage, a good design choice.

The previous method produces sensible results provided that the ripple voltage is between 5% and 20% of the total voltage (it is unusual to allow ripple voltage to rise above this limit).

The effect of ripple voltage on output voltage

The reservoir capacitor charges to the voltage peaks leaving the rectifier, so the ripple voltage is subtracted from this and reduces the output voltage. The output voltage V_{out} can be considered to be made up of two components: V_{DC} which is pure DC, and v_{ripple} which is the superimposed AC ripple voltage. The significance of making this distinction is that subsequent filtering blocks the AC component to leave only the DC component.

$$V_{out} = V_{DC} + v_{ripple}$$

The AC ripple voltage swings symmetrically about V_{DC}, and at its positive peak reaches V_{peak}, therefore:

$$V_{DC} = V_{peak} - \frac{v_{ripple}}{2}$$

Considering our previous example, which had $v_{ripple} = 18\,V$ and $V_{peak} = 325\,V$, the DC voltage that would be seen after subsequent perfect AC filtering would be:

$$V_{DC} = 325\,V - \frac{18\,V}{2} = 316\,V$$

Summarizing, V_{DC} is always reduced by a factor of half the ripple voltage.

Ripple current and conduction angle

Now that we have looked at ripple voltage, we need to look at ripple current. This is the current required by the capacitor to fully recharge it every half cycle. To do this, we need to find the conduction angle, which is the time for which diodes are switched on and the capacitor is charged. See Fig. 5.8.

To do this, we work backwards from the time that the capacitor is fully charged. We know the ripple voltage, so we can find the absolute voltage on

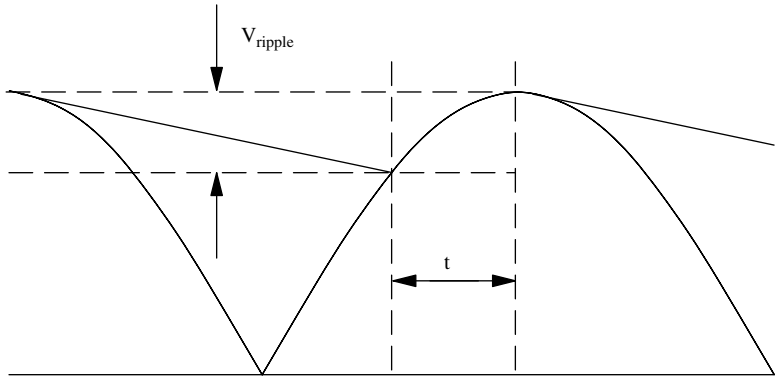


Fig. 5.8 *Determination of conduction angle from ripple voltage*

the capacitor at the instant that the diodes switch on. The voltage at the output of the rectifier (ignoring the polarity) is:

$$v = V_{peak} \cos \omega t$$

At the instant that the diodes switch on, the capacitor voltage must be:

$$V_{peak} - v_{ripple} = V_{peak} \cos(\omega t)$$

Rearranging:

$$\frac{V_{peak} - v_{ripple}}{V_{peak}} = \cos(\omega t)$$

$$\omega t = \cos^{-1}\left(\frac{V_{peak} - v_{ripple}}{V_{peak}}\right)$$

$$t = \frac{1}{2\pi f} \cdot \cos^{-1}\left(\frac{V_{peak} - v_{ripple}}{V_{peak}}\right)$$

If we now put some figures into this equation from our earlier example (Fig. 5.6), *remembering to work in radians and not degrees*:

$$t = \frac{1}{2 \times \pi \times 50} \cdot \cos^{-1}\left(\frac{340 - 18}{340}\right)$$
$$= 1 \text{ ms}$$

The capacitor only draws current from the mains transformer for 1 ms in every 10 ms, or 10% of the time. We should therefore expect this ripple current to consist of short, high current pulses. See Fig. 5.9.

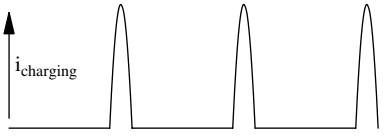


Fig. 5.9 *Ripple current waveform*

We can now find the ripple current using the relationship:

$$i = C\frac{dV}{dt}$$

But we need an expression for $\frac{dV}{dt}$, so we start with our original expression:

$$v = V_{\text{peak}} \cos(\omega t)$$

Differentiating:

$$\frac{dV}{dt} = -\omega V_{\text{peak}} \sin(\omega t)$$

And substituting:

$$i_{\text{ripple}} = -\omega C V_{\text{peak}} \sin(\omega t)$$

If we now put some values into this equation:

$$\begin{aligned} i_{\text{ripple}} &= 2 \times \pi \times 50 \times 68 \times 10^{-6} \times 340 \\ &\quad \times \sin(2 \times \pi \times 50 \times 1 \times 10^{-3}) \\ &= 2.2 \text{ A} \end{aligned}$$

Which is considerably greater than the 120 mA load current!

Quick check: charge is equal to current multiplied by time, which would be *area* on a graph of current against time. If the capacitor has to charge in a tenth of the time that it took to discharge, then it is reasonable to suppose that it will require ten times the current ($Q = It$). This brings us to 1.2 A. However, we observed earlier that the shape of the charging pulse is not rectangular, and because the area under this pulse is smaller than a rectangle of equivalent height and width, this accounts for the final difference in the two answers.

Summarizing: the answer is unexpectedly large, but believable.

The model predicts the maximum possible ripple current, so it is worth ensuring that the rectifier and capacitor could cope with these current pulses by checking their *peak* current rating. In practice, peak ripple current is reduced by:

- Series resistance made up of: diode resistance, capacitor ESR, wiring resistance, transformer winding resistance (secondary and reflected primary).
- Transformer core saturation.

As a result of these factors, ripple current peaks typically range between four and six times the DC load current. As a practical example, a transformer with a silicon bridge rectifier and capacitor input filter producing 108 V_{DC} loaded by a resistor drawing 35 mA_{DC} drew $I_{ripple(pk)} = 160$ mA, a ratio of 4.6:1.

Valve rectifiers have far higher internal resistance than silicon, and may even need additional series resistance in deference to their limited ripple current ratings, so their ratio of I_{ripple}/I_{DC} is likely to be even lower. A 50 MHz Tektronix TCP202 current probe was used to investigate a 300 V HT power supply using a GZ34 to feed its 47 µF polypropylene reservoir capacitor. See Fig. 5.10.

The ripple current pulses contain harmonics of 100 Hz that theoretically extend into low RF. FFT mode was selected on the oscilloscope, allowing spectrum analysis of the reservoir capacitor ripple current. See Fig. 5.11.

The spectrum sweeps linearly from DC (left) to 1.25 kHz (right), so the dominant 100 Hz fundamental can be clearly seen, followed by a train of harmonics. Although the linear vertical scale of this FFT implies that the harmonics die away rapidly, a messier logarithmic scaling revealed that harmonics at 2.5 kHz were only 45 dB below those at 100 Hz.

Transformer core saturation

Toroidal transformers are more susceptible to core saturation as a direct result of their more nearly perfect design. Whether mains or audio, power transformer cores are normally made of GOSS (Grain Oriented Silicon Steel), which has the advantage of allowing a higher flux density in the direction of the grain. Traditional E, I cores are unable to take full advantage of this, since there is always a region where the flux is at right angles to the grain. Toroids, however, have all of their flux aligned with the grain, and can operate at a flux density much closer to saturation, so this is the main reason for the reduced size of toroids. (Operating at higher flux density results permits a smaller core.) Consequently, toroids saturate sharply, whereas E, I cores have a much gentler limit.

Transformer core saturation is undesirable because it releases a leakage field of magnetic flux to induce currents in nearby circuitry. Even worse, this saturation happens cyclically (100 Hz or 120 Hz) and so produces *bursts* of interference with harmonics extending to radio frequencies. Sharper saturation produces a greater proportion of higher harmonics in the leakage field.

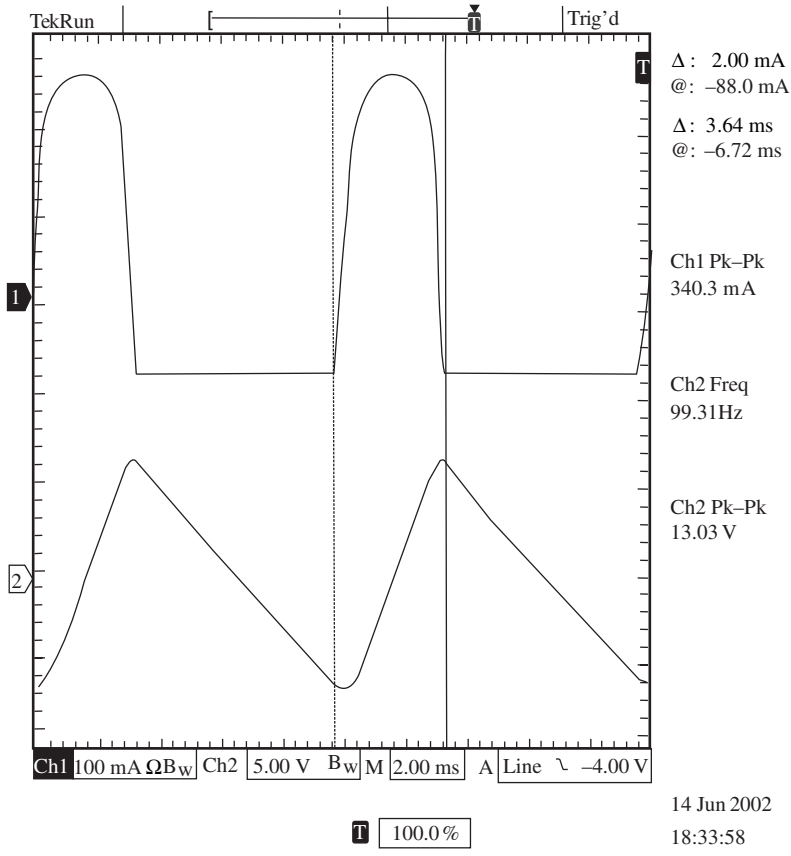


Fig. 5.10 *Captured reservoir capacitor waveforms due to 88 mA load current. Upper trace (Ch1): current waveform ($I_{pk} = 340$ mA). Lower trace (Ch2): ripple voltage (13 V_{pk-pk})*

This is not merely an apocryphal tale of woe. The author tore his hair out searching for the source of (video) hum in a picture monitor, only to find that the cause was a saturating mains toroid inducing hum directly into the neck of the picture tube.

Choosing the reservoir capacitor and transformer

If we have designed our supply to have a ripple voltage of 5% of supply voltage, then for 90% of the time, the transformer is disconnected, and the output resistance of the power supply is determined purely by the capacitor

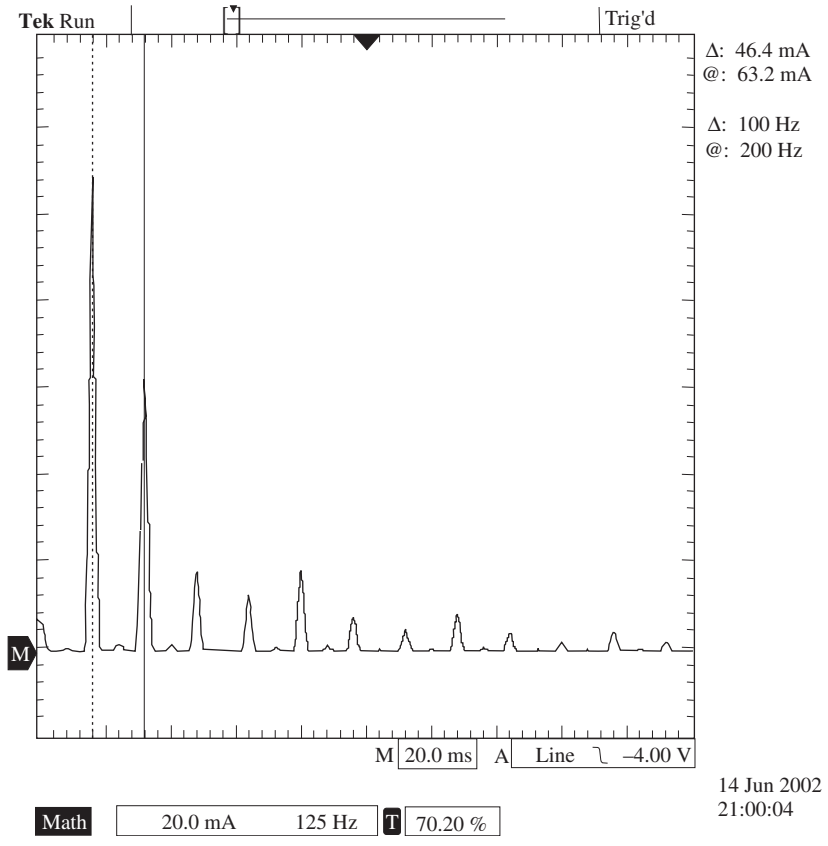


Fig. 5.11 *Spectrum of reservoir capacitor ripple current*

ESR and associated output wiring resistance. This is why changing reservoir capacitors from general purpose types to high ripple current types, produces a noticeable effect on the sound of an amplifier, they have a lower ESR (but a higher price).

The transformer/rectifier/capacitor combination is a non-linear system. This makes its behaviour considerably more complex than the ideal Thévenin source, so we need to investigate it over different periods of time.

In the short term (less than one charging cycle) the output resistance of the supply is equal to capacitor ESR plus wiring resistances. This will be true even for very high current transient demands, which may appear in each and every charging cycle, provided that they do not significantly change the charge on the capacitor. All that is required is that the capacitor should be able to

source these transient currents. To be able to do this, the capacitor needs a low ESR, not just at mains frequencies, but also up to at least 40 kHz, because a Class B power output stage causes a rectified (and therefore frequency doubled) version of the audio signal to appear on the power supply rails. (See Chapter 6 for explanation of Class B.) We can cope with this requirement by using an electrolytic capacitor designed for use in switch mode power supplies as the main reservoir, and bypass it with smaller capacitors. See Fig. 5.12.

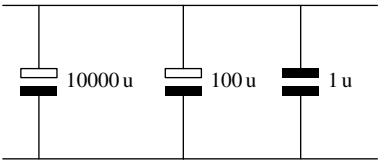


Fig. 5.12 *Use of bypass capacitors to simulate 'perfect' capacitor*

A power amplifier may significantly deplete the charge in the reservoir capacitor, causing output voltage to fall either by drawing a sustained high current, due to a continuous full power sine wave test, or by reproducing a short, but loud, sound – such as a bass drum.

Supplying a constant load is relatively easy, because we know exactly how much current will be drawn, and we simply design for that current. If the ripple voltage for a sensible ripple current is higher than we would like, then we simply add a regulator to remove it.

The difficulties start when we want to supply a changing load. It might seem that if the power amplifier is rated at 100 W continuous into 8 Ω, then all we have to do is to calculate what load current that implies, and design for that current. The drawbacks of this approach are more easily demonstrated using a transistor amplifier, where the load is directly coupled to the output stage and power supply:

$$P = I^2 R$$

Therefore, for a sine wave:

$$i_{\mathrm{RMS}} = \sqrt{\frac{W}{R}} = \sqrt{\frac{100}{8}} \approx 3.5\,\mathrm{A}$$

But we have to supply the *peak* current, which is $\sqrt{2}$ greater, at 5 A. The amplifier might need ±50 V rails, so a power supply capable of delivering

±50 V at 5 A is implied. We therefore need a 500 W power supply to feed a 100 W (mono) amplifier!

This is a very large and expensive power supply, and we would need some very good reasons for using it.

The key to the problem lies in the class of the output stage. (See Chapter 6.) If the output stage operates in pure Class A, then the quiescent current equals the peak current required at maximum power output, in this case 5 A. If we really draw a constant 5 A from the power supply, then we genuinely do need a 500 W power supply. See Fig. 5.13.

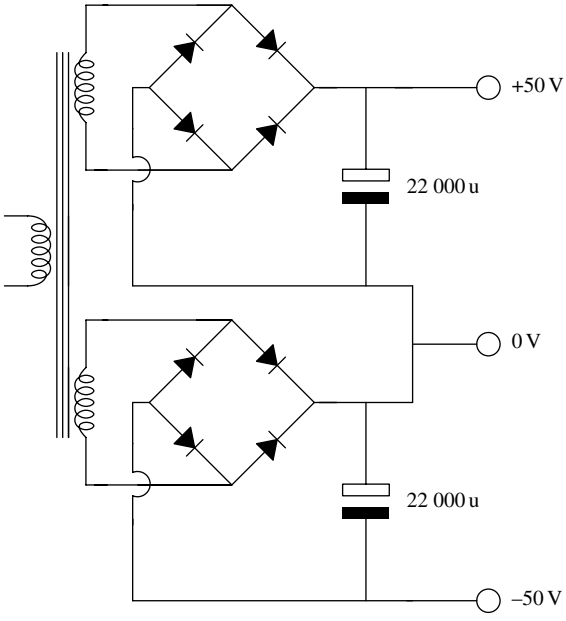


Fig. 5.13 *Typical power supply for transistor amplifier*

The reservoir capacitor value was easy to determine using our earlier formula and 5% ripple voltage criterion, but the transformer is quite a different matter. It is possible to determine the requirements of the transformer exactly, using the graphs originally devised by O.H. Schade.[4] In practice, the required transformer information may not be available, so a practical rule of thumb is to make the VA rating of the transformer at least equal to the required output power.

If the amplifier output stage is Class B, then it still supplies 5 A to the load on the crests of the sine wave, but at other points in the cycle, the required

current from the power supply is much lower. The effect of the reservoir capacitor is to average the fluctuating current demand, and for a sine wave:

$$I_{\text{average}} = 0.637 I_{\text{peak}}$$

The average supply current is 3.2 A, so a 350 VA transformer would be chosen.

We could further argue that the amplifier does not operate at full power all the time, and that the short-term musical peaks requiring maximum output power do not last long. A smaller transformer could therefore be used, since the reservoir capacitor could supply the peak currents. This is a very seductive argument, and many commercial amplifier manufacturers have been persuaded by it, since £1 extra on component cost generally adds £4–£5 to the retail price.

We do not have to work to such tight commercial considerations and, within reason, the bigger the mains transformer, the better.

The choke input power supply

Choke input power supplies were very popular in the heyday of valve amplifiers for the simple reason that large value capacitors were not available, so chokes had to be used for smoothing. See Fig. 5.14.

If we were able to make a choke input supply with a choke of infinite size, the mains transformer current would be identical to the DC load current.

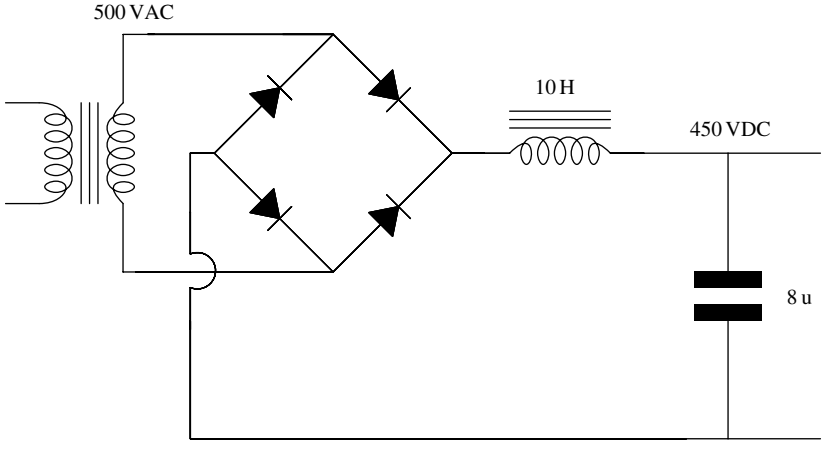


Fig. 5.14 *Choke input power supply*

Practical supplies do not quite achieve this ideal, so the transformer current is a combination of DC load current and a somewhat smaller, nearly sinusoidal current drawn by the choke. Nevertheless, the choke input power supply has the great advantage that it draws a very nearly continuous current from the mains transformer rather than a series of high current pulses. To understand why this is so, we need to consider the output waveform of the rectifier in detail. See Fig. 5.15.

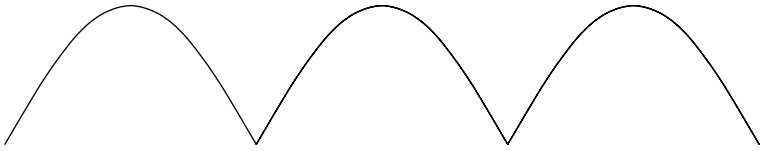


Fig. 5.15 *Full-wave rectified AC sine wave*

This waveform is a full-wave rectified sine wave, but because it has undergone a non-linear process (rectification), the frequencies present in this waveform are not the same as went into the rectifier. Fourier analysis reveals that the result of full-wave rectification of a sine wave is:

$$V = \frac{2\sqrt{2}v_{\text{in(RMS)}}}{\pi}\left(1 + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cos 2n\omega t}{4n^2 - 1}\right)$$

Note that $v_{\text{in(RMS)}}$ is the voltage *before* rectification.

The previous equation is a mathematical way of expressing an infinite series, but for our purposes, it is much easier to present the information as follows:

$$V = v_{\text{in(RMS)}}[0.90 + 0.6(2f) - 0.12(4f) + 0.05(6f) - 0.03(8f)\ldots]$$

This shows us that a full-wave rectified sine wave is made up of a DC component corresponding to $0.90v_{\text{in(RMS)}}$, plus a series of decaying even harmonics of the input frequency (f) *before* rectification. The choke has such a high reactance to these AC terms that only the DC component reaches the load. The output voltage of a choke input power supply is therefore $0.90v_{\text{in(RMS)}}$, rather than $\sqrt{2}v_{\text{in(RMS)}}$ for the capacitor input supply.

Minimum load current for a choke input supply

Unfortunately, a minimum load current must be drawn from a choke input power supply before it operates correctly. If less than this current is drawn, the circuit reverts to pulse charging of the capacitor, and the output voltage rises to a maximum of $\sqrt{2}v_{\text{in(RMS)}}$. The absolute minimum current that should be drawn is:

$$I_{\text{min.}} = \frac{2\sqrt{2}v_{\text{in(RMS)}}}{6\pi^2 fL}$$

In practice, the inductance of the choke depends partly on the current through it (*BH* curve), so it is wise to draw rather more current than this, and a handy approximation (appropriate for 50 Hz or 60 Hz mains) is:

$$I_{\text{min.(mA)}} \approx \frac{v_{\text{in(RMS)}}}{L_{\text{(H)}}}$$

Choke input supplies invariably feed a capacitor, and the minimum current requirement is therefore important, since insufficient current could cause the voltage across the capacitor to rise to ≈157% of nominal voltage, which might destroy it. The traditional way of dealing with this problem was to use a *swinging* choke, whose inductance rose at low currents, and although these became unfashionable after the 1960s, they have returned – just like flares and platform shoes.

Once the minimum current has been exceeded, the output ripple is constant with load current, and the AC components of the rectified sine wave are attenuated by a factor of:

$$\frac{v_{\text{ripple(RMS)}}}{V_{\text{DC}}} = \frac{1}{6\sqrt{2} \cdot \omega^2 LC}$$

where $\omega = 2\pi f$.

Alternatively, if we consider that only the amplitude of the 2nd harmonic is significant, the equation can be rearranged into the following forms:

$$v_{\text{ripple(RMS)}} \approx \frac{v_{\text{in(RMS)}}}{3.7L_{\text{(H)}}C_{\text{(μF)}}}(50\,\text{Hz}),$$
$$\approx \frac{v_{\text{in(RMS)}}}{5.4L_{\text{(H)}}C_{\text{(μF)}}}(60\,\text{Hz})$$

where L is in H, and C in μF, and $v_{in(RMS)}$ is the mains transformer secondary voltage.

Current rating of the choke

Although an infinitely large choke would allow the choke and mains transformer to have a current rating equal to the maximum DC load current, they actually have to support a somewhat higher current, and it is particularly important that the choke is correctly rated. Remember that the choke generates magnetic flux in a core proportional to the current passing through the coil, but if too much magnetizing force is applied, the core saturates, causing its inductance to fall to zero.

Since the output of the rectifier comprises a DC component and an AC component, it is the summation of these components that determines the current rating of the choke. The DC component is simply the load current, but the AC component requires a little more thought.

Because the choke is followed by a capacitor, which is a short circuit to AC, the entire AC component leaving the rectifier is developed across the reactance of the choke, causing an AC current to flow. Once we know the AC voltage across the choke, we can easily calculate the current.

As previously mentioned, the AC component is dominated by the 2nd harmonic, so we can simplify the calculation to deal exclusively with this component.

The instantaneous AC voltage across the choke is therefore:

$$v = 0.6v_{in(RMS)} \cos(2\pi ft)$$

where f is the 2nd harmonic of mains frequency.

The reactance of the choke is:

$$X_L = 2\pi fL$$

Using Ohm's law to combine the two equations, the instantaneous current through the choke is:

$$i_{AC} = \frac{0.6v_{in(RMS)} \cos(2\pi ft)}{2\pi fL}$$

We are only concerned with the maximum current, which occurs when $\cos(2\pi ft) = 1$, so this factor can be removed, leaving:

$$i_{AC(peak)} = \frac{0.6v_{in(RMS)}}{2\pi fL} \text{ (2nd harmonic only)}$$

It was stated that only the 2nd harmonic was significant, but this assumption should now be examined. Referring to the Fourier series, the 4th harmonic is 20% (0.12/0.6) of the voltage of the 2nd. The doubled reactance of the choke at the 4th harmonic halves the choke current, resulting in a 4th harmonic current that is only 10% of the 2nd, so the approximation is fair, but there is room for improvement.

The sum of the AC currents drawn by each of the Fourier terms, up to and including the 8th harmonic, was investigated graphically on a computer to find the maximum positive peak. (The negative peak is irrelevant since when added to the DC load current, it reduces the total peak current.) The result of this exercise modified the equation to:

$$\begin{aligned} i_{AC(\text{positive peak})} &= \frac{0.544v_{in(RMS)}}{2\pi fL} = \frac{v_{in(RMS)}}{1155L} (50\,\text{Hz}) \\ &= \frac{v_{in(RMS)}}{1386L} (60\,\text{Hz}) \end{aligned}$$

But the total peak current flowing through the choke is the sum of the AC peak current and the DC load current:

$$i_{\text{total peak current}} = I_{DC} + i_{AC(peak)}$$

As an example, a Class A power amplifier using a pair of push–pull 845 valves requires a raw HT of 1100 V at 218 mA, and a 10 H 350 mA choke is available, but is this adequate? The transformer supplying the choke input filter has an output voltage of $1224v_{RMS}$. Using the previously derived equation and assuming 50 Hz mains:

$$i_{AC(\text{positive peak})} = \frac{v_{in(RMS)}}{1155L} = \frac{1224}{1155 \times 10} = 106\,\text{mA}$$

$$\begin{aligned} i_{\text{total peak current}} &= I_{DC} + i_{AC(peak)} = 218\,\text{mA} + 106\,\text{mA} \\ &= 324\,\text{mA} \end{aligned}$$

The total peak current is 324 mA, so the 350 mA rated choke is just sufficient, but the example shows that choke AC current can be surprisingly high, particularly when high HT voltages are contemplated.

The best way of finding a suitable choke for a choke input supply is to write the various choke equations into a spreadsheet and test available chokes for suitability. As a sweeping statement, chokes for HT choke input supplies generally need to be $\geq$15 H, otherwise the AC current becomes crippling.

Transformer current rating for a choke input supply

The peak choke current must be supplied by the transformer, so the transformer should be rated appropriately. However, since transformer ratings assume resistive loads and sine waves, their current ratings are RMS of sine wave, and they can deliver a peak current of $\sqrt{2}$ this value, so the previous example would require a transformer with an RMS sine wave current rating of 229 mA (324 mA_{pk}). This is sufficiently close (5% error) to the DC load current of 218 mA that a common approximation is to assume that the transformer should have an AC_{RMS} current rating equal to the DC load current.

Current spikes and snubbers

Choke input power supplies are not perfect, and have two main problems: electrical switching spikes and mechanical vibration.

Although we said earlier that the choke input power supply drew a continuous current from the mains transformer, this cannot be exactly true. Since the rectifier diodes require a certain voltage across them before they switch on (irrespective of whether they are thermionic or semiconductor), there must be a time, as the input waveform crosses through zero volts, when neither diode is switched on. The current drawn from the transformer is therefore not quite continuous, and must momentarily fall to zero. The choke will try to maintain current, and in doing so will develop a voltage:

$$V = -L\frac{di}{dt}$$

In any full-wave rectifier, the diodes switch off at twice mains frequency, and at that instant, $di/dt = \infty$, so theoretically infinite voltage spikes are produced with a repetition rate of twice mains frequency. See Fig. 5.16.

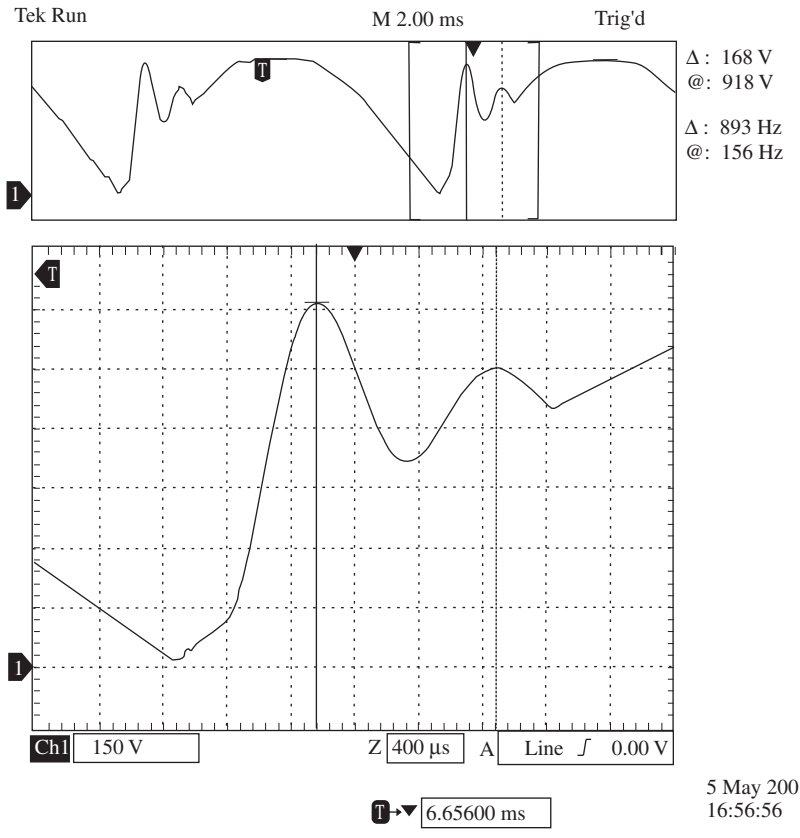


Fig. 5.16 *Extreme choke ringing caused by rectifier switching without load current*

Although drawing a significant load current greatly damps the ringing of the choke, the current waveform still has a glitch. See Fig. 5.17.

Traditionally, a resistor/capacitor snubber network was connected across the choke to protect the interwinding insulation of the mains transformer from the spikes. See Fig. 5.18a.

Although fitting the traditional 10 nF + 10 kΩ snubber across the choke tames the voltage spikes, it degrades high frequency filtering and worsens the glitch in the current waveform. See Fig. 5.19.

A snubbing method that significantly improves high frequency filtering is to fit back to back capacitors across the choke, with their centre tap connected to 0 V, and use the internal resistance of the choke as the snubbing resistance. Optimum high frequency filtering is obtained by choosing C_1 so that it

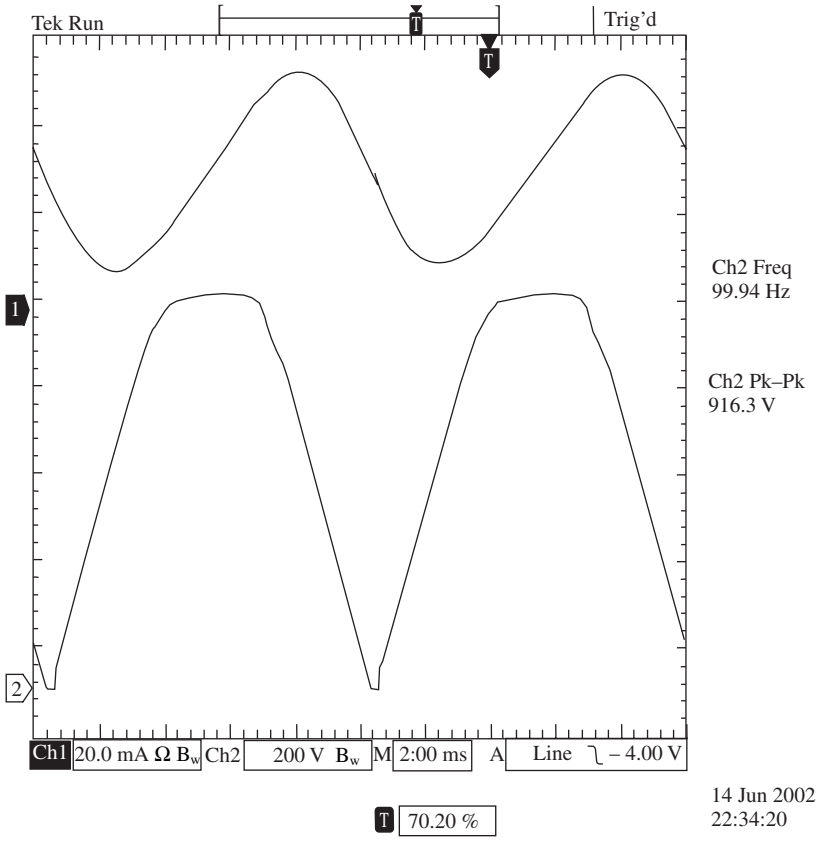


Fig. 5.17 *Without snubber. Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier*

resonates with the leakage inductance of the mains transformer at the same frequency as the self-resonance of the choke, but this seems not to be critical, and curiously, 220 nF is often a practical value for both HT and LT supplies. See Fig. 5.18b.

The modified snubber network removes the voltage spikes without compromising HF filtering or adding glitches to the current waveform. See Fig. 5.20.

As mentioned previously, the entire AC component at the output of the rectifier is across the choke. In Chapter 4 we observed that output transformers could 'sing' due to loose laminations or magnetostriction, and the same is true here. The choke could buzz at twice mains frequency, and if it has any loose parts, such as a loose screening can, it will rattle. Loudly. Even worse,

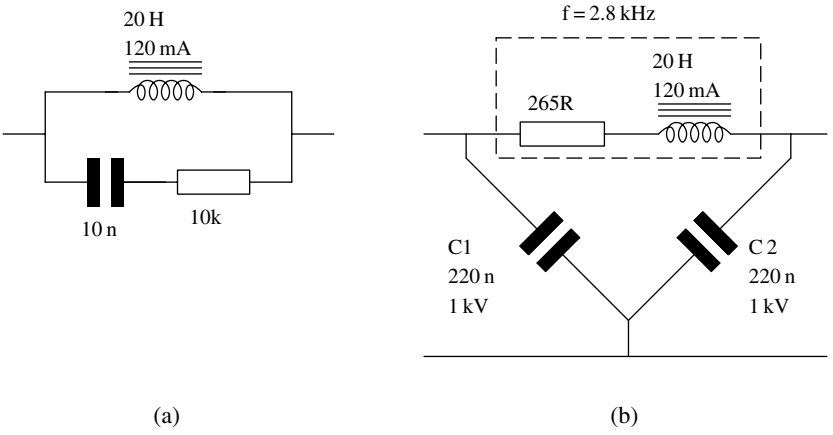


Fig. 5.18 *Traditional and improved choke snubber networks*

the choke is bolted to a resonant sounding board (the chassis) which will amplify the buzz.

The author has recently investigated a number of choke input power supplies. A buzzing choke implies core saturation. Unfortunately, it seems that iron cores may significantly deteriorate over the years, reducing inductance, which increases the AC current, perhaps to the point of saturation, thus causing buzz. If you must use old chokes, check them for buzz under load *before* drilling holes in the chassis.

Using small reservoir capacitors to reduce HT voltage

For a given input voltage, a choke input supply produces the lowest output voltage ($0.90v_{in(RMS)}$) because only the DC component from the rectifier reaches the load, whereas a capacitor input supply with $C_{reservoir} = \infty$ achieves the maximum ($\sqrt{2}v_{in(RMS)}$) because it can use the AC component. Another way of looking at a choke input supply is to consider it to be a capacitor input supply where $C_{reservoir} = 0$. We can now see that changing the reservoir capacitor value could be a useful way of adjusting output voltage between $0.90v_{in(RMS)}$ and $\sqrt{2}v_{in(RMS)}$, thus allowing a previously unsuitable transformer secondary to provide the required output voltage without wasting power in a resistor. A secondary advantage of this approach is that the very low ripple current drawn by a small reservoir capacitor greatly reduces the generation of RF noise.

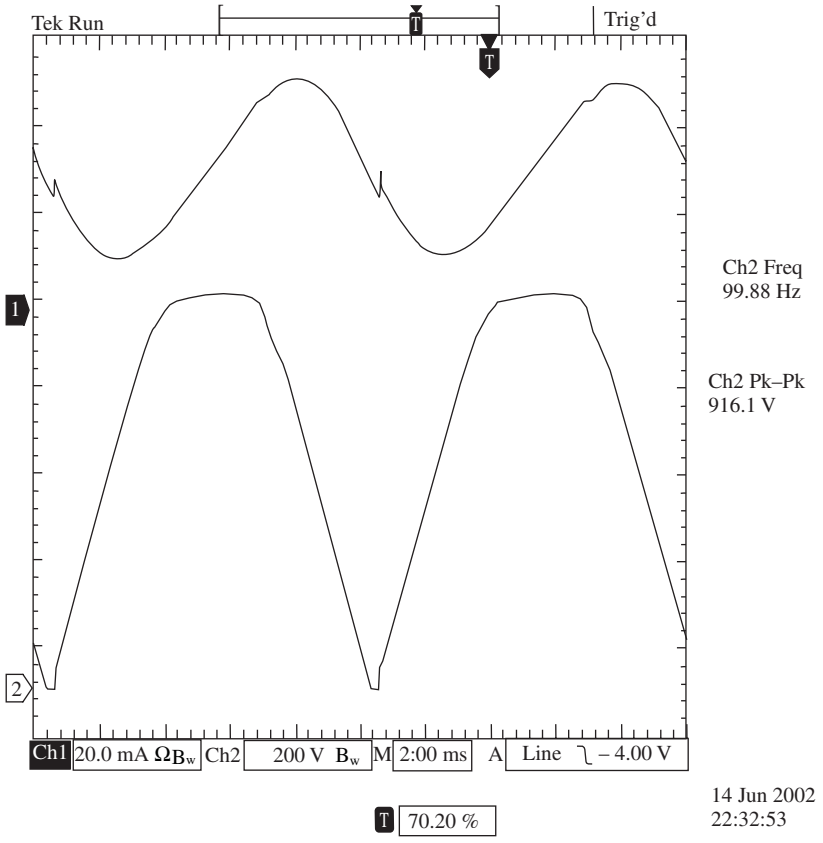


Fig. 5.19 *Choke input supply with 10 nF +10 kΩ snubber. Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier. Note the worsened current waveform*

Alternatively, when a traditional valve rectifier is used for the positive HT, it requires a centre-tapped transformer, but that same winding could also provide a negative rail, perhaps for grid bias of output valves or constant current sink tails. Using valve rectifiers for negative rails is awkward because it requires a pair of individual rectifiers such as EY84, and to avoid exceeding $V_{hk(max.)}$, they usually need a dedicated heater supply. Silicon rectifiers are the obvious choice, and although they are slightly noisier than EY84, the low ripple current demanded by the small reservoir capacitor reduces their noise.

It would be nice to have a simple equation for determining the exact HT voltage from a reservoir capacitor size, but we contravene the fundamental

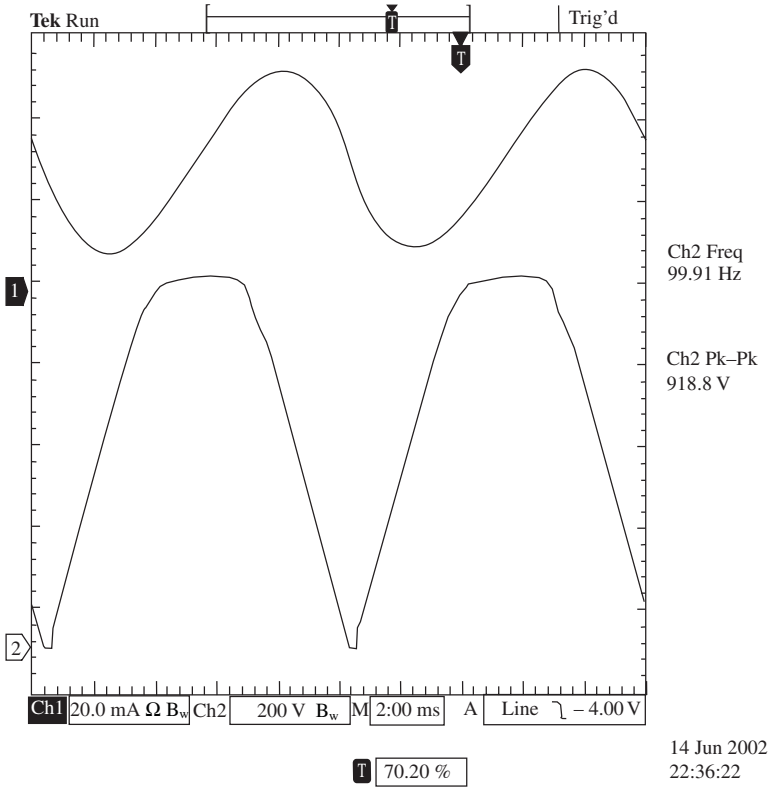


Fig. 5.20 *With 220 nF back to back snubber. Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier. Note the complete absence of glitches*

assumptions made for capacitor input smoothing, and do not consider the DC component leaving the rectifier. The capacitor input equations are a good starting point, but the final value of reservoir capacitor has to be adjusted on test.

Broad-band response of practical *LC filters*

So far, our investigation of rectification and filtering has focused on low frequency behaviour, but we now need to broaden our outlook to include behaviour from DC to low radio frequencies. To attenuate low (≈100 Hz)

frequencies significantly, an *LC* filter with a large inductance is required, but this inevitably has internal shunt capacitance. Conversely, the capacitor has series inductance, and these hidden components mean that any practical *LC* filter has a frequency response that may be divided into four main regions. (Although surprisingly smooth, the graph is a result of practical measurements of an *LC* filter.) See Fig. 5.21.

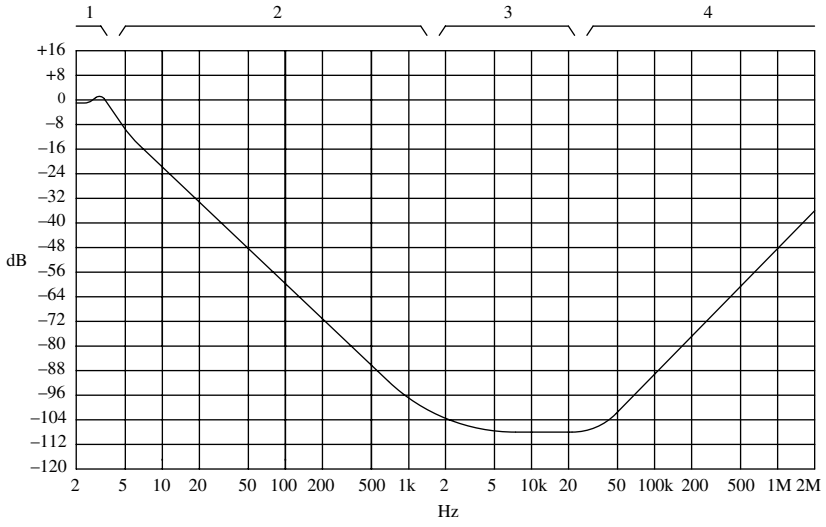


Fig. 5.21 *Measured frequency response of LC filter (20 H 50 mA, 120 μF 400 V polypropylene)*

Region 1

This is the only region we can directly control, so it is well worthy of investigation. Apart from losses due to DC resistance, the low-pass filter does not attenuate frequencies below the LF resonance:

$$f_{\text{res.(LF)}} = \frac{1}{2\pi\sqrt{LC}}$$

We aim to position the (hopefully) subsonic resonance as low as possible by making *L* and C large, because every octave by which we can lower $f_{\text{res.(LF)}}$

produces an additional 12 dB of filtering. If $f_{res.(LF)}$ has $Q > 0.707$, an LF peak results in the response of the filter, so it is useful to check Q:

$$Q = \frac{1}{R_{DC}} \cdot \sqrt{\frac{L}{C}}$$

where: L = inductance of choke
R_{DC} = internal resistance of choke
C = capacitance of smoothing capacitor

Ideally, the resonance should be critically damped ($Q = 0.5$), which can be achieved by adding external series resistance to the choke. Strictly, the load resistance across the capacitor also damps the resonance, and this may be transformed into a notional extra choke series resistance using:

$$r_{notional} = \frac{L}{CR_{load}}$$

However, the damping effect of the load resistance is usually negligible. (A series regulator is a constant current, or infinite, AC load to the smoothing circuitry, so it adds no damping whatsoever.)

As a typical traditional example: a filter uses a 15 H choke with 260 Ω internal resistance coupled to an 8 μF paper/oil capacitor, resulting in $f_{res.(LF)} = 14.5$ Hz, $Q = 5.27$. This Q is too large, and $f_{res.(LF)}$ is too near the audio band, but the additional 2.48 kΩ series resistance required to achieve critical damping would waste HT voltage and greatly increase power supply output resistance. A better approach would be to replace the 8 μF capacitor with a 120 μF polypropylene, since this would give $f_{res.(LF)} = 3.75$ Hz, $Q = 1.36$, and this Q might be acceptable. Adding 447 Ω of additional series resistance would achieve $Q = 0.5$.

Region 2

The reactance of the choke doubles for each octave rise in frequency, whilst the reactance of the reservoir capacitor halves, producing the familiar 12 dB/octave slope.

Region 3

The shunt capacitance of the choke begins to take effect. Once the reactance of the shunt capacitance is equal to the inductive reactance of the choke, the choke resonates, so this region may be defined as beginning at $f_{res.(HF)}$.

Above this self-resonant frequency (3–15 kHz for a typical HT choke), the shunt capacitance forms a potential divider with the smoothing capacitor whose loss is constant with frequency.

$$\text{Loss(dB)} \approx 20 \log \left(\frac{C_{\text{choke}}}{C_{\text{smoothing}}} \right)$$

Region 4

The series inductance of the reservoir capacitor becomes significant, and this forms a hidden high-pass filter in conjunction with the shunt capacitance of the choke, so the output noise of the practical filter rises at 12 dB/octave.

These filtering concepts can be simplified by considering an idealized *LC* filter response to be made up of three straight lines having freedom to move either vertically or horizontally. See Fig. 5.22.

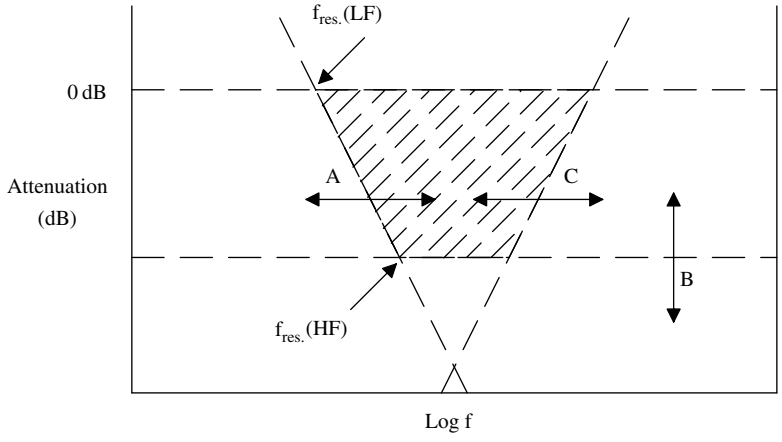


Fig. 5.22 *Conceptual model of universal LC filter*

- Line A falls at 12 dB/octave, and slides horizontally to the left as choke inductance and smoothing capacitance increase.
- Line B falls vertically as choke stray capacitance falls. (Capacitance between adjacent winding layers of the choke could be reduced by interposing earthed electrostatic screens.)
- Line C rises at 12 dB/octave, and slides horizontally to the right as smoothing capacitor series inductance falls. (Minimize the length of wire from capacitor foils to junction with choke and load.)

An optimum *LC* filter maximizes the hatched area. There is no point in reducing choke stray capacitance further once line B meets the intersection of lines A and C.

Wide-band filtering

Lines A and C in the idealized model show that any *LC* filter can only be effective over a limited range of octaves, and that its filtering must deteriorate at high frequencies. Thus, wide-band filtering requires a cascade of different filters positioned over different octave ranges. We filter low frequencies first because:

- The output of the rectifier contains large amplitudes at low frequencies, for which an LF choke is designed, but which would saturate an RF choke.
- As frequency increases, stray inductance and capacitance become more significant. This means that even 100 mm of wire has inductance, and could be used as a choke at UHF, but it is also an aerial whose length should be minimized near the load.

Typically, we might start with a 15 H choke input supply using a 120 μF filter capacitor to reduce ripple to less than a volt. A typical 15 H choke is likely to have $f_{\text{res.(HF)}} \approx 3$ kHz, so if we wanted to continue our 12 dB/octave filtering indefinitely, we would arrange for our next *LC* filter to take over from the first at the intersection between lines A and B by setting its $f_{\text{res.(LF)}} \approx 3$ kHz.

$$f_{\text{res.(LF)}} = \frac{1}{2\pi\sqrt{LC}}$$

So:

$$LC = \frac{1}{(2\pi f)^2} = \frac{1}{(2 \times 3.14 \times 3000)^2} = 2.8 \times 10^{-9}$$

This *LC* product could be achieved by a 100 mH choke wound on a ferrite core in conjunction with a 33 nF capacitor connected close to the load.

If fitted, any regulator should be connected after the second stage of filtering, in order to minimize its exposure to RF. In critical situations, a third stage might use a VHF ferrite bead connecting to the anode load

resistor via a feedthrough capacitor soldered to the circuit's screening can, directly adjacent to the anode load, thus ensuring that the final high frequency filtering is performed as close to the load as possible, and minimizing the length of aerial.

Sectioned *RC* filters

We might have carefully designed an HT power supply (choke or capacitor input) so that it can be achieved with the parts available, so that it produces $2\,V_{pk-pk}$ of ripple. However, we might need $<1\,mV_{pk-pk}$, but can afford to drop some voltage. We need a filter that will attenuate the ripple by a factor of >2000. Since an RC filter is a potential divider, the attenuation is $=R/X_c$ (provided that this ratio is reasonably large). Suppose that in our example, we can tolerate $2\,k\Omega$ of resistance, so $X_c = 2\,k\Omega/2000 = 1\,\Omega$. Since the ripple frequency is 100 Hz, we find the required capacitance using:

$$C = \frac{1}{2\pi f X_c} = \frac{1}{2 \times \pi \times 100 \times 1} = 1590\,\mu F$$

This is a very large capacitor, and represents a brute force solution to the problem. A more efficient alternative is to make a filter out of a cascade of sections each using a smaller resistor and capacitor. See Fig. 5.23.

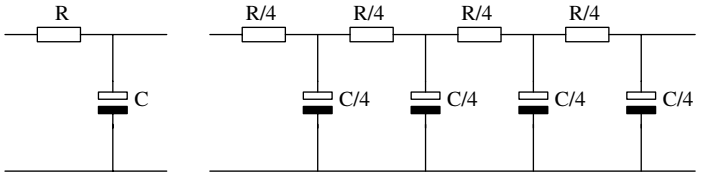


Fig. 5.23 *Sectioning the RC filter leaves total resistance and capacitance unchanged, but increases attenuation by changing the ultimate slope from 6 dB/octave to 24 dB/octave*

The problem is to determine how many sections are ideal. Fortunately, Scroggie[5] (writing as 'Cathode Ray') has already investigated this problem and produced a handy table:

No. of sections	$2\pi fCR$ (R_{total}/X_c)	Attenuation	RC in kΩ · μF per section	
			100 Hz	120 Hz
1	16	16	25.5	21.2
2	45.6	130	18.1	15.1
3	90	997	15.9	13.3
4	149	7520	14.8	12.4
5	223	56 400	14.2	11.8
6	311	420 000	13.8	11.5

(This table differs slightly from the original reference because Scroggie did not have the benefit of a spreadsheet when calculating his values.)

To understand the table using our example: we need attenuation >2000, so the first number of sections that can exceed this in the attenuation column is $n = 4$. If the total resistance is to be 2 kΩ, then each section must be 2 kΩ/4 = 500 Ω. To find the individual capacitance required, we use the 100 Hz RC column. The capacitance required is 14.8/0.5 = 29.6 μF. In practice, we would probably use 470 Ω resistors and 33 μF capacitors. The key point is not simply that the four 33 μF capacitors are likely to be much cheaper (and smaller) than a single ≈1590 μF capacitor, but that the sectioned filter promises almost four times the attenuation.

Alternatively, you have a huge bag of 22 μF capacitors, and enough room to use four, but must use 2.5 kΩ of series resistance. How can the capacitors be best used to attenuate 100 Hz ripple? Connecting the four capacitors in parallel gives a total capacitance of 88 μF, so the ratio of $R_{total}/X_c = 138$. Inspecting the R_{total}/X_c column, to find the first number of sections that exceeds 110, we see that three sections should be used. Each resistance is thus 2.5 kΩ/3 = 833 Ω. If we only use three sections, our total ratio R/X_c is reduced by a factor of $^3/_4$, so it falls to 104, but this is still optimum with three sections and gives an attenuation of 997, whereas using four 22 μF capacitors in parallel with a 2.5 kΩ series resistor would only have given an attenuation of 138. Our factor of improvement is seven, yet we have used one capacitor less.

Voltage multipliers

We have previously looked at rectification and smoothing with the aim of producing a DC supply capable of sourcing significant current with minimum

noise. Occasionally, we need a DC potential, but are not too worried about current capability or noise. Typical examples are the final EHT required by the display tubes in televisions (10–25 kV) and analogue oscilloscopes ($\approx$10 kV), or the polarizing bias required by electrostatic loudspeakers ($\approx$5 kV).

Voltage multipliers were originally needed by physicists to develop the 800 kV accelerating potentials needed to investigate whether the impact of hydrogen ions on matter would produce X-rays. The Cockcroft–Walton[6] ladder, or multiplier, can be extended indefinitely, with each step theoretically adding $\sqrt{2}V_{\text{in(RMS)}}$ to the output, but its regulation is very poor. Each diode needs a voltage rating $> \sqrt{2}v_{\text{in(RMS)}}$. Unfortunately, all but the lowest capacitor must be rated at $> 2\sqrt{2}v_{\text{in(RMS)}}$. Additionally, because succeeding capacitors are charged by rectifier switching that partly discharges the lowest capacitor, this capacitor must be of a higher value to reduce voltage drop. See Fig. 5.24.

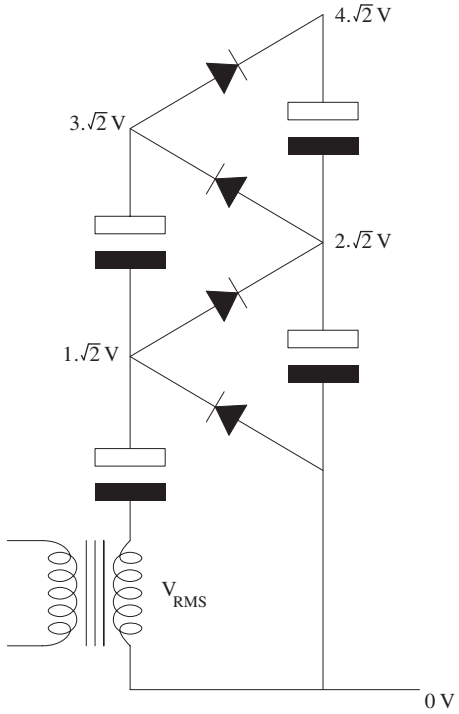


Fig. 5.24 *Cockcroft–Walton voltage multiplier*

Although voltage multipliers were invented for generating EHT, they can be useful for providing negative grid bias, and the 6 W Rogers Cadet stereo power amplifier even used a voltage doubler for its main HT. There are two forms. See Fig. 5.25.

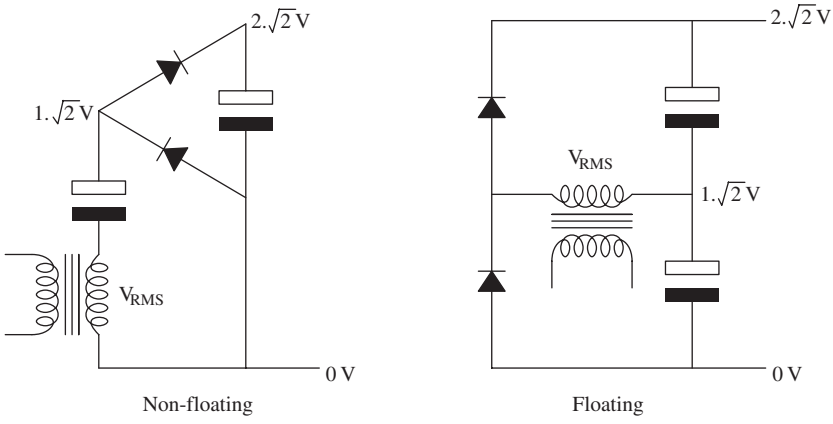


Fig. 5.25 *Non-floating and floating voltage doublers*

The non-floating doubler is a truncated Cockcroft–Walton ladder. It can be connected in parallel with a conventional centre-tapped rectifier/transformer combination allowing a subsidiary (higher voltage) HT to be developed – perhaps for polarizing a dedicated high frequency electrostatic loudspeaker.

The advantage of the floating doubler is that it uses two identical capacitors, each rated at half the output voltage, but the diodes must be rated at $>2\sqrt{2}v_{in(RMS)}$. Because each capacitor is only charged on alternate half cycles, ripple voltage is doubled compared to a conventional full-wave rectifier. Because the ripple voltages of the two capacitors are in series, a further doubling of ripple voltage occurs. Thus, for a given ripple voltage, the floating doubler requires each capacitor to have four times the capacitance needed by a conventional full-wave rectifier.

Regulators

The best way of improving a power supply is to use a voltage regulator. A voltage regulator is a real world approximation of a Thévenin source; it has a fixed output voltage, and an output resistance that approaches zero. A true Thévenin source implies infinite current capacity, whereas the supply that

feeds a regulator has limited current capacity. It is therefore important to realize that the regulator can only simulate a Thévenin source over a limited range of operation, so we must ensure that we remain within this range under *all* possible operating conditions.

All voltage regulators are based on the potential divider. Either the upper or the lower leg of the divider is made controllable in some way, and by this means, the output voltage can be varied. See Fig. 5.26.

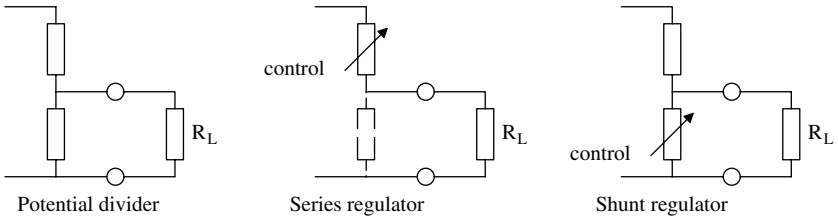


Fig. 5.26 *Relationship between voltage regulators and the potential divider*

If the upper element is made controllable, then the regulator is known as a *series* regulator, because the controlled element is in *series* with the load. If the lower leg of the divider is controlled, then the regulator is known as a *shunt* regulator, because this element is *shunted* by the load. Shunt regulators are usually inefficient compared to series regulators, and their design has to be carefully tailored to their load, but they have the advantage that they can both source and *sink* current.

The fundamental series regulator

The fundamental elements of a series voltage regulator are shown in Fig. 5.27.

This circuit is shown using semiconductors, but a valve version could equally well be built. The error amplifier amplifies the difference between the reference voltage and a fraction of the output voltage, and controls the series pass transistor such that a stable output voltage is achieved.

The circuit depends for its operation on negative feedback. We saw in Chapter 1 that when feedback is applied, input and output resistances change by the feedback factor $(1 + \beta A_0)$. Voltage regulators rely on reducing the output resistance of the system by the ratio of the feedback factor.

Suppose initially that the regulator is working, and that there is 10 V at the output. By potential divider action, there must be 5 V on the inverting input of the operational amplifier. The voltage reference is holding the non-inverting input at 5 V. The series pass transistor is an emitter follower fed by the error amplifier, and has 10 V on its emitter, so the base must be at 10.7 V.

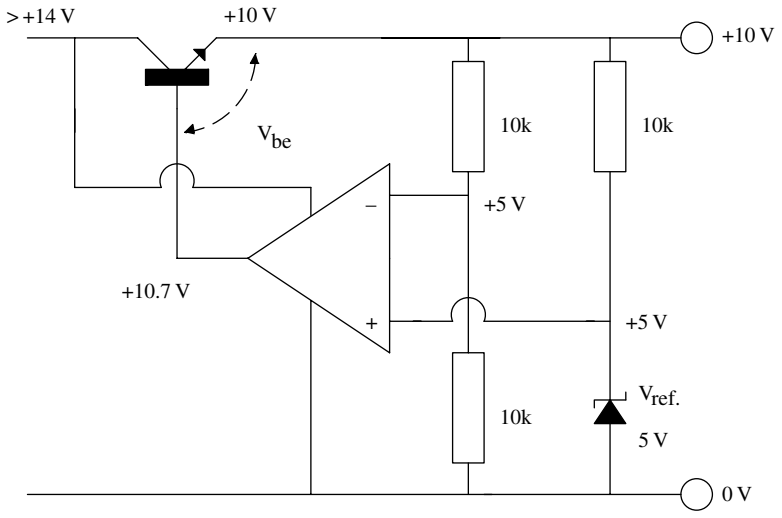


Fig. 5.27 *Series regulator*

Suppose now that the output voltage falls for some reason. The voltage at the midpoint of the potential divider now falls, but the voltage reference maintains 5 V. The error amplifier now has a higher voltage on its non-inverting input than on its inverting input, and its output voltage must rise. If the voltage on the base of the transistor rises, its emitter voltage must also rise. The circuit therefore opposes the reduction in output voltage.

Since the same argument works in reverse for a rise in output voltage, it follows that the circuit is stable and that the output voltage is determined by the combination of the potential divider and the reference voltage. If we redraw the regulator, we can easily see that it is simply an amplifier whose gain is set by the potential divider, and that it amplifies the reference voltage. See Fig. 5.28.

By inspection, the output voltage is therefore:

$$V_{out} = \frac{x + y}{y} \cdot V_{ref.}$$

Since the error amplifier simply amplifies the reference voltage, any noise on the reference will also be amplified, and we should feed it from as clean a supply as possible. Although the argument seems like a snake chasing its own tail, if we feed the reference from the output of the supply (which is clean), then the reference will be clean, and the output of the supply will also be

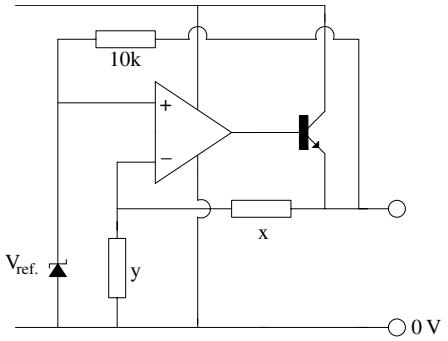


Fig. 5.28 *Series regulator redrawn to show kinship to non-inverting amplifier*

clean. It might be thought that supplying the current for the reference voltage from the output voltage might cause instability, but in practice this is not a problem.

It should be noted that all regulators need an input voltage higher than their output voltage. The minimum allowable difference between these voltages before the regulator fails to operate correctly is known as the drop-out voltage (because the regulator 'drops out' of regulation). With this particular design, it is only a few volts, but drop-out voltage for a valve version could be 40 V or more.

The two-transistor series regulator

The two-transistor series regulator is a very common and useful circuit. See Fig. 5.29.

This circuit is often used because of its extreme cheapness, but despite that, its performance is really quite good. Q2, the series pass transistor, is fed from the collector of Q1, a common emitter amplifier. The emitter of Q1 is held at a constant voltage by the voltage reference, whilst its base is fed a fraction of the output voltage by the potential divider. If the output voltage rises, Q1 turns on harder, drawing more current, its collector voltage (connected to the base of Q2) falls, causing the emitter voltage of Q2 (which is the output voltage) to fall, thus counteracting the initial error. This circuit is ideal for use as a bias voltage regulator in a power amplifier, because we often need to drop more volts than an IC regulator would tolerate.

As presented, the circuit can only supply 50 mA of output current because the base current for Q2 is stolen from the collector current of Q1. If we

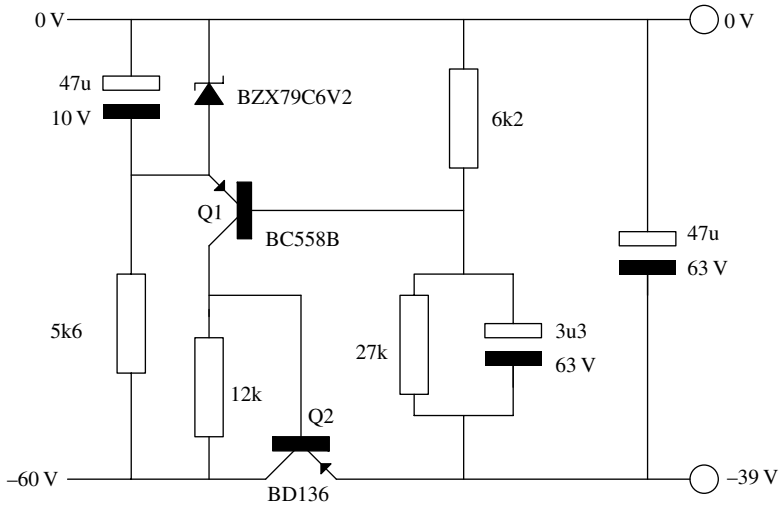


Fig. 5.29 *Basic two-transistor negative regulator*

increased the collector current of Q1, Q2 could steal more, and output current could be increased, but a better solution would be to replace Q2 with a Darlington transistor, which would need less base current. Alternatively, Q2 could be a power MOSFET, but it would need a gate-stopper resistor of ≈100 Ω soldered directly to its gate pin.

The Zener diode passes 12 mA, which is quite sufficient to ensure that it operates correctly, and has a stable output voltage with minimum noise. A 6.2 V Zener has been chosen because it has lowest temperature coefficient and lowest slope resistance, but it still produces some noise, so it is bypassed by the 47 μF capacitor.

The *speed-up* capacitor[7]

This capacitor is connected across the upper resistor of the potential divider. Its purpose is to increase the amount of negative feedback available at AC, and thereby reduce hum and noise. Since any linear regulator can be considered to be composed of an op-amp enclosed by a feedback loop, a generic graph may be drawn. See Fig. 5.30a.

The op-amp gain is a combination of DC open loop gain and gain that falls at 6 dB/octave with frequency. The gain within the cross-hatched

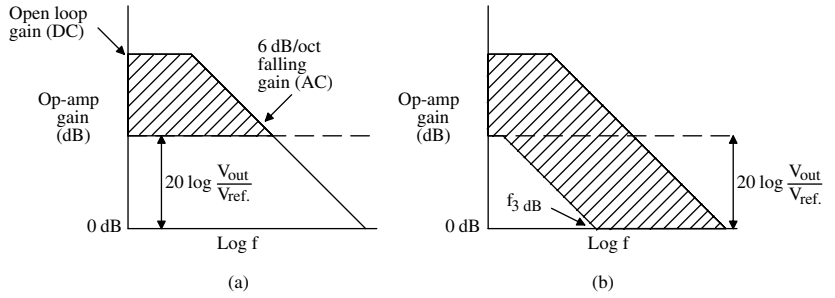


Fig. 5.30 *The effect of the speed-up capacitor on ripple rejection*

area is available for attenuating incoming ripple, so ripple reduction is maximized by:

- Maximizing DC open loop gain.
- Maximizing high corner frequency (741: $f_{corner} \approx 20\,Hz$, 5534: $f_{corner} \approx 1\,kHz$).
- Minimizing the ratio of DC output voltage to reference voltage.

Although we need the op-amp to have the correct gain at DC to set the output voltage correctly, all of the gain below the cross-hatched area is being wasted. The *speed-up* capacitor aims to recover this wasted gain. See Fig. 5.30b.

At first sight, it would seem that $f_{3\,dB}$ should be placed sufficiently low that all of the previously wasted gain is recovered. However, an oversize speed-up capacitor slugs the response of the regulator to changes in load current.

The maximum value for this capacitor is found by first calculating the AC Thévenin resistance that it sees:

$$r_{\text{Thévenin}} = \frac{1}{\frac{1}{h_{ie}} + \frac{1}{x} + \frac{1}{y}}$$

Remembering that:

$$h_{ie} = \frac{h_{fe}}{gm}$$

And also that:

$$gm = 35I_c$$

We find that for this circuit, with $h_{fe} = 200$, and $I_c = 12\,\text{mA}$, $h_{ie} \approx 500\,\Omega$. So the Thévenin resistance seen by the capacitor is $\approx 450\,\Omega$.

We would like the capacitor to have a significant effect on the lowest ripple frequency to be attenuated, which is 100 Hz (120 Hz USA). The potential divider chain and capacitor is a step equalizer whose effect on the regulator is similar to that used for the RIAA 3180 μs/318 μs pairing in Chapter 6. We could make the reactance of the capacitor at the lowest ripple frequency equal to the Thévenin resistance at the tapping of the potential divider, which would mean that an infinitely large capacitor could only improve ripple reduction by a further 3 dB:

$$C = \frac{1}{2 \times \pi \times 100 \times 450}$$
$$= 3.5\,\mu\text{F}$$

The nearest value is 3.3 μF. This is quite a small capacitor, and the author has seen many similar circuits with oversized capacitors, and, indeed, built one himself. The subjective effect of the oversize capacitor was to create a bass boom that was incorrectly thought to be due to room acoustics.

At the opposite limit, we could set the reactance of the capacitor relative to the total resistance of the divider chain. The smaller capacitor would only give a 3 dB improvement in hum compared to a chain without a capacitor, but its LF transient response would be better than a regulator using a larger capacitor.

The value of the 'speed-up' capacitor is a compromise between hum reduction and regulator LF transient response, so there isn't a 'correct' answer here, other than that the capacitor should be small. You might even want to determine its final value by listening because different loudspeakers (with different LF damping) can prefer different values.

Compensating for regulator output inductance

The regulator also has a capacitor across its output. As shown in Fig. 5.30, the gain of the error amplifier falls with frequency due to Miller effect and stray capacitances, so the amount of gain available for reducing output impedance falls. If $(1 + \beta A_0)$ has fallen, then the output impedance must rise, and the effect is that output impedance rises with frequency. A perfect Thévenin source in series with an inductor would look identical, and for this reason the output of regulators is often described as being inductive at high frequencies. The shunt capacitor maintains a low output impedance at high frequencies.

A variable bias voltage regulator

We often need a bias voltage regulator to be variable between certain limits. In this example, we will look at a grid bias regulator needed for an 845 directly heated triode. Perusal of RCA anode characteristics (circa 1933) indicated a grid bias voltage of −125 V, but modern valves do not match the original curves exactly, and we must equalize anode currents in this (push–pull) output stage to avoid saturating the output transformer with an unbalanced DC current and causing distortion. A range of ±25 V either side of the nominal −125 V seems reasonable, but how do we design a regulator to fulfil this requirement?

Fortunately, since it supplies a part of the circuit where the signal voltages are very high (up to 90 V_{RMS}), the regulator need not necessarily have an impeccable noise performance, and Zener diodes are perfectly acceptable. See Fig. 5.31.

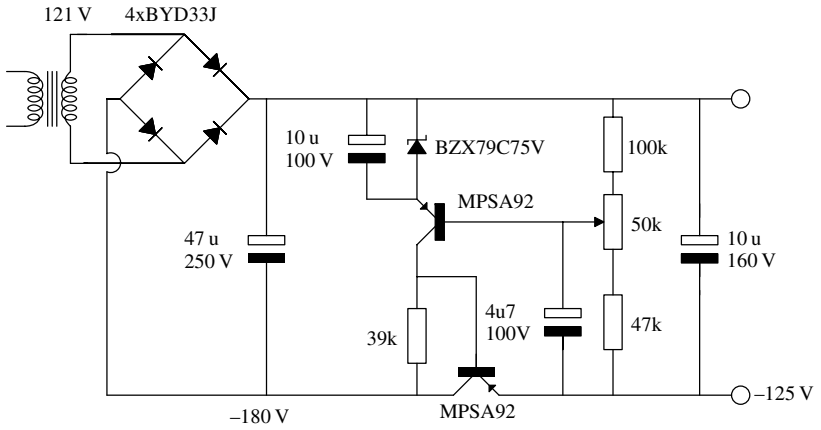


Fig. 5.31 *Adjustable −125 V bias regulator*

A higher voltage Zener allows the finished circuit to have better regulation, but we must still allow a reasonable voltage between the collector and emitter of the control transistor. In practice, a Zener voltage of about half the maximum output voltage is usually a good choice, and 75 V Zeners are readily available.

The Zener holds the emitter of the transistor at −75 V, and $V_{be} = 0.7$ V, so the base of the transistor will be held at a fixed potential of −75.7 V. Since the base of the transistor is connected to the wiper of the potential divider, the wiper must also be held at −75.7 V, no matter what output

voltage is set. We can now calculate the required attenuation of the potential divider for the two extreme design cases:

$$\frac{100\,\mathrm{V}}{75\,\mathrm{V}} = 1.323$$

$$\frac{150\,\mathrm{V}}{75.7\,\mathrm{V}} = 1.9815$$

By choosing a convenient value for the variable resistor in the middle of the potential divider, we now have enough information to calculate the required resistors either side. A low value of variable resistor would require a large current to flow in the potential divider, whereas too high a value will cause errors due to the (small) base current drawn by the transistor. A good engineering principle is that the potential divider chain should pass roughly ten times the expected base current, so a 50 kΩ variable resistor was a convenient standard value for this example.

When the wiper of the variable resistor is set to produce the largest voltage at the regulator output, it is connected directly to the grounded resistor (x), and vice versa. Using the standard potential divider equation, for −150 V:

$$\frac{x + y + 50}{x} = 1.9815$$

And similarly for −100 V:

$$\frac{x + y + 50}{x + 50} = 1.323$$

We now have two equations that can be solved, either simultaneously, or by substitution, to give the values of the fixed resistors x and y. In this particular case, the values fell out very conveniently to give $x = 100\,\mathrm{k\Omega}$ and $y = 47\,\mathrm{k\Omega}$, where x is the upper potential divider resistor and y is the lower.

The 317 IC voltage regulator

Although the two-transistor regulator is the ideal choice for a bias regulator, because of its high voltage drop capability, once we need higher currents at lower voltages, its limitations become apparent.

It is perfectly possible to build a voltage regulator using a handful of components including an operational amplifier, a voltage reference, various transistors, resistors and capacitors. With care, the circuit can be made to work almost as well as an IC regulator, and only cost about three times as much. We need not feel guilty about using IC regulators.

The 317 is a standard device that is made by all the major IC manufacturers.[8] Linear Technology[9] makes an upgraded version of the 317, the LT317, but the only difference is that the guaranteed tolerance of the voltage reference is tighter. A commercial design could therefore set its output voltage using fixed resistors rather than a variable resistor, thus saving money, because not only are variable resistors expensive to buy, but they also have to be adjusted (which costs money). We do not often have to worry about such considerations, so the standard 317 is fine.

The 317 incorporates all of the fundamental elements of a series regulator in one three-terminal package, and we need only add an external potential divider to produce an adjustable regulator. See Fig. 5.32.

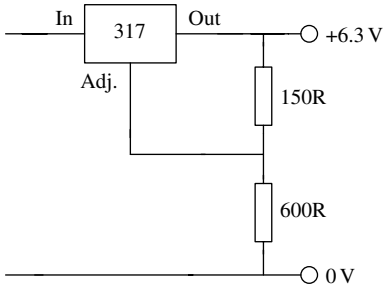


Fig. 5.32 *Basic 317 regulator circuit*

One end of the voltage reference is connected to the OUT terminal whilst the other is an input to the error amplifier. The other input of the error amplifier is the ADJ terminal. The 317 therefore strives to maintain a voltage equal to its reference voltage (1.25 V) between the OUT and ADJ terminals. All we have to do is to set our potential divider so that the voltage at the tap is $V_{out} - 1.25\,V$, and the 317 will do the rest.

In data sheets for the 317, you will invariably find that the upper resistor of the potential divider is 240 Ω. The reason for this is that the 317 must pass 5 mA before it can regulate reliably. If the potential divider passes 5 mA, then this ensures that the device is able to regulate even if there is no external load.

The 317 sources ≈50 μA of bias current to the opposite rail from the ADJ pin, which therefore flows down the lower leg of the potential divider. Normally, this is negligible, but if you are designing a high voltage regulator, and choose a lower potential divider current, this will need to be taken into account.

The manufacturers' data sheets generally show a regulator with the ADJ pin bypassed to ground by a 10 μF electrolytic, which improves ripple rejection from 60 dB to 80 dB at 100 Hz. This is directly equivalent to the speed-up capacitor that we added in the two-transistor regulator, but because the reference voltage is tied to V_{out}, rather than ground, the 'speed-up' capacitor connects to ground, rather than V_{out}.

We could therefore use the method derived earlier to check that this is the optimum value of capacitor. The ADJ pin is an input to an operational amplifier, so we can treat it as infinite input resistance, and we are only concerned with the external resistor values. If we were to use an upper resistor of 240 Ω, and a 3.9 kΩ lower resistor to set an output voltage of 22 V, then the optimum value of capacitor would be 7 μF, so a 10 μF electrolytic is a reasonable choice, although the author would probably prefer 6.8 μF if he had one in stock.

Just like the two-transistor regulator, the output of the 317 is inductive, and the manufacturer's output impedance curves suggest that the output impedance is equivalent to ≈2.2 μH in series with 2.7 μΩ, so they recommend a 1 μF tantalum bead output bypass capacitor, as shown in the equivalent circuit. See Fig. 5.33.

If we assume that a tantalum bead is a perfect capacitor (!), then we have an underdamped resonant circuit, and we can calculate its Q:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}} = \frac{1}{2.7 \times 10^{-6}} \cdot \sqrt{\frac{2.2 \times 10^{-6}}{1 \times 10^{-6}}} = 550\ 000$$

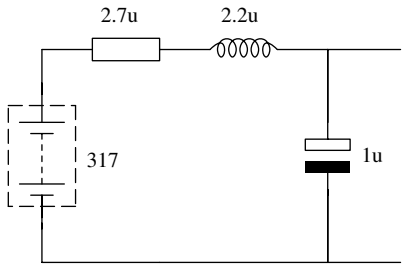


Fig. 5.33 *AC Thévenin equivalent of 317 plus 1 μF bypass capacitor*

Stray resistance will reduce this Q considerably, but it will not reduce it to $Q = 0.5$, which would be critically damped. This would not matter greatly, because we would be unable to excite the circuit from the output (any external excitation would be short circuited by the capacitor). If we now concede that the capacitor is not perfect, we may be unlucky enough to be able to excite the resonance, and the circuit could become unstable. Rearranging the formula, $3\,\Omega$ critically damps the resonance, so the manufacturers recommend $2.7\,\Omega$ in series with the tantalum capacitor.

LT supplies and common-mode noise

Classic pre-amplifiers had AC heaters and suffered hum. Modern pre-amplifiers have DC heaters, but because this high current (1–2 A) is awkward to smooth adequately purely passively, they invariably add a regulator to reduce hum to mV levels. The fact that the heater supply is now stabilized is a bonus.

Almost every effort within the *design* of audio circuitry to increase immunity to HT noise increases sensitivity to LT noise because it increases r_k (see Chapter 2), but heater supplies are somewhat different from HT supplies, and cause considerable confusion. LT supply ripple may be defined as differential-mode noise, since it is the difference in voltage between one heater pin and another. Provided that differential-mode noise is reasonably low, the considerable thermal inertia of an indirectly heated cathode filters it out *completely*, and differential RF noise is irrelevant.

Nevertheless, there has been a trend to using exotic (and expensive) heater regulators in the belief that this allows the finished amplifier to sound better. If valve heaters were designed to be supplied by $6.3\,V_{RMS}$ of AC ripple, it hardly seems likely that they will be materially affected by minor differences between heater regulators producing perhaps $<10\,mV_{pk-pk}$ ripple, since this is essentially perfection anyway, so why are there believable reports of audible improvements caused by changing heater regulators?

Valves may not be particularly sensitive to differential-mode noise on LT supplies, but they certainly are sensitive to common-mode noise. In the context of an LT supply, common-mode noise leaves the potential difference between the ends of the heater constant, but *both* voltages are bouncing up and down, and if we view the heater as a single conductor, its voltage with respect to the cathode is changing, rendering it perfectly capable of inducing noise into the cathode.

Common-mode heater noise is a problem for small-signal valves because the noise current is capacitively coupled from the heater directly to the

enclosing cathode. Unless decoupled by a cathode capacitor, the noise current develops a noise voltage at the cathode primarily determined by r_k, which is then added to the wanted signal and amplified by the valve. Cathode followers have a low (but non-zero) r_k, and differential pairs have quite a high r_k, so it is clear that we need to be able to prevent common-mode noise from reaching the heater supply.

Because regulators are designed to address differential-mode noise they are normally ineffective against common-mode noise, although their support circuitry may inadvertently assist in rejecting common-mode noise. Useful common-mode filtering is only gained by employing RF techniques involving series RF chokes and shunt capacitors to chassis. See Fig. 5.34.

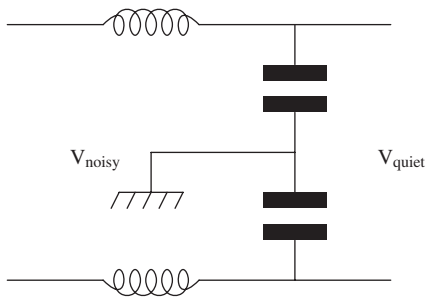


Fig. 5.34 *Filtering common-mode noise*

Because the heater is capacitively coupled to the cathode, and the reactance of C_{h-k} is low at RF, the cathode is likely to be particularly sensitive to RF for two reasons:

- In a single-ended stage, the cathode will be (should be) decoupled to ground using a large capacitor. But the capacitor is not connected directly onto the cathode emissive surface, so inductance in the connecting wires reduces its effectiveness at RF. Additionally, the (inevitably electrolytic) capacitor itself has some inductance and ESR.
- In a differential pair, the cathode unavoidably has quite a high resistance to ground (via the anode load resistors – not the cathode resistance), and cannot form a useful *CR* filter in conjunction with C_{hk}. We are forced to rely on the (usually quite poor) RF balance of the differential pair to reject RF, so a differential pair is likely to be more sensitive to heater-borne RF than a single-ended stage.

Sources of common-mode LT noise

Common-mode noise currents can be capacitively coupled to the LT winding from any adjacent winding on the transformer, with high voltage windings being the worst offenders. Although there is often an electrostatic screen between the mains winding and the nearest secondary, it is most unusual to find a screen around the HT winding. This is significant because HT rectifier diodes generate copious RF noise when they switch, which is easily coupled via the interwinding capacitance. As an example, one version of the Bevois Valley amplifier (see Chapter 6), used a Leak Stereo 20 as the donor chassis (retaining the GZ34 rectifier and transformers), but HT switching spikes were immediately observed on the $6.3\,V_{AC}$ LT supply when the amplifier was first tested.

Typical mains transformers have an interwinding capacitance of ≈1 nF between adjacent windings, so the HT supply should ideally use a separate transformer. However, the problem may be alleviated if this is not possible. The first step in the cure is to connect a small (≈10 nF is fine) capacitor from each side of the heater supply to chassis using very short leads. The two capacitors are in parallel so far as common-mode noise is concerned, so their 20 nF forms a potential divider in combination with the 1 nF interwinding capacitance. See Fig. 5.35.

Because the reactance of a capacitor is inversely proportional to its capacitance, the equation for the loss through this potential divider is found using the *other* capacitor value as the numerator:

$$\frac{V_{out}}{V_{in}} = \frac{1n}{20n + 1n} = 0.048 = -26\,\text{dB}$$

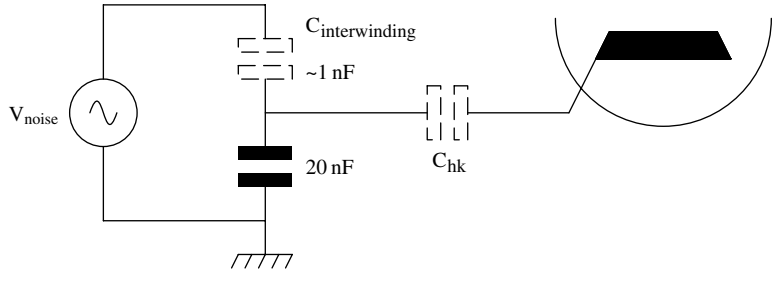


Fig. 5.35 *Potential divider formed by interwinding capacitance and heater bypass capacitors*

Notice that a potential divider formed of only one type of component (*R or* C *or L*) attenuates all frequencies equally, so RF noise at *all* frequencies is now attenuated by 26 dB, but we can do better than this. If we were to add equal inductance to each leg of the heater supply, the resulting *LC* filter would improve attenuation. Since we are trying to filter common mode noise, rather than differential mode, we can wind a bifilar choke on a small ferrite core without worrying about core saturation, because the currents in the two coils create equal and opposite fields, which cancel, so the core does not see any net magnetization.

The 317 as an HT regulator

Because the 317 is a floating regulator, there is no reason why it should not be used to regulate a 400 V HT supply. However, because the 317 can only tolerate 37 V from input to output, it needs support circuitry for protection.[10] See Fig. 5.36.

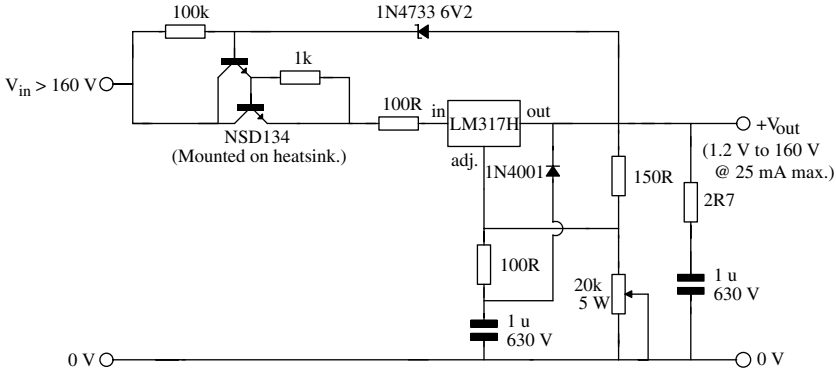


Fig. 5.36 *Basic high voltage regulator (reprinted by kind permission of National Semiconductor)*

The 317 is preceded by a high voltage transistor Darlington pair whose sole aim in life is to maintain 6.2 V between the IN and OUT terminals of the 317, longevity of the 317 is thus ensured. The Darlington pair can easily cope with variations in mains voltage, but it should not be thought that this circuit is proof against a short circuit when used at typical valve voltages.

Accidentally short circuiting a regulator of this type with an oscilloscope probe results in an almighty bang and the silicon is destroyed. The author *knows*.

The lower arm of the potential divider is bypassed, but has a resistor in series with the capacitor to improve LF transient response by raising the lower $f_{-3\ dB}$ frequency of the step equalizer, and a diode has been added, allegedly to discharge the capacitor in the event of an output short circuit (although the author's experience is that it doesn't actually help).

This fundamental circuit will be investigated further when we design a complete HT and LT power supply.

Valve voltage regulators

Valve voltage regulators have always been very rare, and we will now see why. See Fig. 5.37.

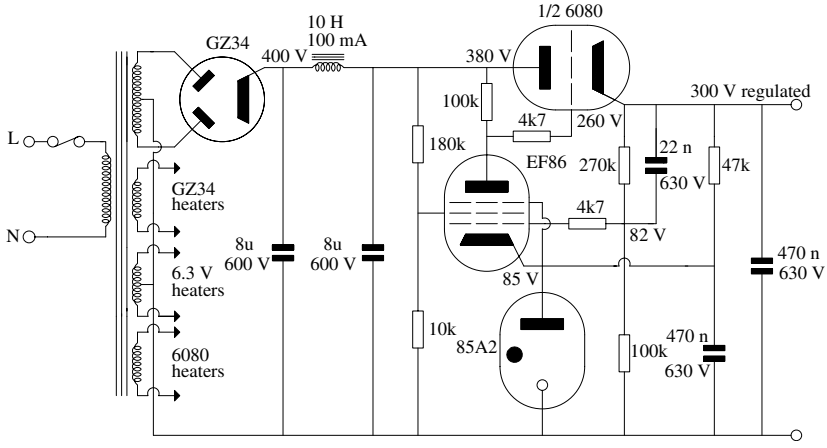


Fig. 5.37 *Basic valve voltage regulator*

The circuit is very similar to the two-transistor regulator; it simply has valves and higher voltages. The Zener diode has been replaced by a neon stabilizer which *burns* at 85 V, this holds the cathode of the EF86 stable, and the grid is fed from the potential divider. The series pass element is a 6080 double triode ($P_{a(max.)} = 13\,W$) which was specifically designed for use in series regulators and can pass high currents at low anode voltages.

A valve rectifier is used, and in deference to its limited ripple current capacity, an 8 μF paper/foil capacitor has been chosen for the reservoir, although a polypropylene capacitor ($\leq 60\,\mu F$ for this rectifier) would be physically practical. This results in considerable ripple voltage which is filtered by the following *LC* combination.

Unless the g_2 resistor value is carefully chosen, the performance of the regulator is slightly compromised by feeding g_2 of the EF86 from the raw supply, but if fed from the regulated supply, there is a danger that the circuit might not switch on, and might simply sulk. The gain of the EF86 is ≈100, and above ≈100 Hz this gain is available for reducing the output resistance of the 6080, whose $gm \approx 7\,\text{mA/V}$, so $r_k \approx 200\,\Omega$ (including the effect of the external $100\,\Omega$ R_a). The regulator therefore achieves an output resistance of $\approx 2\,\Omega$. The EF86 is somewhat noisy (2 μV), but this noise is swamped by the noise from the 85A2, which produces 60 μV.

Even reference valves such as the 85A2 are notorious for *voltage jumps*, an effect whereby the DC reference voltage jumps by typically 5 mV if the operating current is changed. Maximum stability is thus achieved by stabilizing the valve at the manufacturer's preferred operating current, but if this current is changed, even if it is returned to its original value, the valve takes time to recover its original stability. Thus, these valves can be considered to be 'one use' devices, and should not be salvaged from existing equipment if maximum stability is needed. Fortunately, there are huge stocks of NOS 85A2, so it is not a problem to use an unused valve for each new circuit.

Because each valve has its cathode floating above ground, three separate heater supplies are needed (the 85A2 is a cold cathode valve). The EF86 could be supplied by a grounded heater supply, but this is putting quite a strain on the heater to cathode insulation. The 6080 data sheet specifies $V_{hk(max.)} = 300\,\text{V}$, so if $V_{regulated} < 300\,\text{V}$, the 6080 heater could also be powered from a grounded heater supply. Stressing heater/cathode insulation reduces the life expectancy of valves and increases noise, so it is not recommended.

The author once built a rather more complex 420 V valve regulator with floating heater supplies and an error amplifier made of two cascaded ECC81 differential pairs. This had a commensurately lower output resistance, and the measured hum and noise was $400\,\mu\text{V}_{pk-pk}$. It was big. It was heavy. It drifted. Viewed on an AC coupled oscilloscope, the output voltage gently perambulated around 420 V due to heater voltage variation (unregulated AC heater supplies). The 317 regulator is rock steady.

Optimized valve voltage regulators

Oscilloscope design presents many challenges because a bandwidth of DC to at least 20 MHz is required. Valve oscilloscopes required stable, quiet HT supplies, so their voltage regulators were carefully optimized.[7] Heater

voltages were stabilized against mains voltage variation by control circuits involving a saturable inductor in series with the mains winding of the heater transformer.[11]

Any voltage regulator can be improved by increasing the gain of its error amplifier. A single triode has the lowest gain, but a pentode (or cascode) has higher gain. If even greater gain is required, a pair of stages can be cascaded (more than two stages would be impractical because phase shifts would almost certainly turn the regulator into a power oscillator). Because the error amplifier amplifies DC, drift must be minimized, so the first stage of a high gain regulator *must* be a differential pair, and a dual triode is convenient. The second stage is much more flexible, and could be another triode differential pair, or a single-ended stage using either a triode or a pentode.

Using a pentode's g_2 as an input for hum cancellation

If a pentode is used as the second stage, g_2 can be considered to be an inverting input. If the correct proportion of raw HT ripple is injected at this point, it will cancel at the anode, resulting in a regulator with no hum at its output, but the tactic is not without its problems:

- For the pentode to operate correctly, g_2 must be at the correct DC potential. This is usually derived from a potential divider across the (clean) output of the supply. A large value resistor can then be connected from the raw HT, and its value adjusted until ripple is cancelled. The exact value of this resistor is awkward to calculate because we do not usually know the value of $\mu_{(g_2-a)}$, so its value is usually determined by experiment. Values could be anywhere in the range from 150 kΩ to 1.5 MΩ.
- Although valve manufacturers specified most parameters quite tightly, we now rely on an unspecified parameter, and there is no guarantee that valves made by different manufacturers that meet all the specified parameters will match our unspecified parameter. As an example of this problem, the four-tube EMI2001 colour camera set beam current by controlling g_2, so when a new tube was ordered (£1500 a shot in 1986), it was necessary to specify that the tube was to be used in a 2001. Similarly, Tektronix stocked selected tubes (valves), not because they were better than any others, but because they were guaranteed to work correctly in *their* circuit.
- Variations between valves mean that the cancellation is not perfect, but any remaining ripple is easily mopped up by the loop gain of the error amplifier.

Increasing output current cheaply

The majority of circuitry within oscilloscope and audio circuitry is Class A, so it draws a very nearly constant current. One of the functions of a regulator is to regulate output voltage against changes in load current, but if the current is almost unchanging, much of the regulating ability is being wasted. As an example, the regulator might face a load with a quiescent current of 100 mA but that could rise to 150 mA, or drop to 50 mA under certain circumstances. We could design the regulator to be able to pass 150 mA, but this would need a bigger series pass valve. Instead, we could bypass the series pass valve with a resistor that allowed 50 mA to flow directly into the load. The series pass valve now only has to pass 100 mA under full load. When the load requires only 50 mA, this is provided entirely by the bypass resistor, and the regulator is in danger of dropping out of regulation, so this condition sets the limit for the maximum current that may be bypassed by a resistor.

Adding the bypass resistor slightly increases ripple because it injects raw HT into the clean circuit, but because the output resistance of the regulator is likely to be $<1\,\Omega$, potential divider action greatly reduces the added ripple.

As an example, the circuit in Fig. 5.38 incorporates both of these modifications.

The regulator showcases some other tricks that improve performance.

As previously mentioned, neon references are noisy, but because we have chosen to use a differential pair, the neon now drives a high impedance input, so we can add a filter to reduce noise. The capacitor previously across the neon reference has been removed because of the danger of it causing oscillation when excited by voltage jumps (this was previously damped by r_k of the valve). Additionally, the current through the neon reference has been stabilized at the preferred operating current, so jumps should be minimal.

The ECC83 differential pair has its anodes at 209 V, and although it would just be possible to direct couple this voltage to the grid of the EF91 pentode, its cathode would be at ≈ 213 V, which would not only cause problems with V_{hk}, but would reduce gain because of the necessarily high value of R_k. To reduce this problem, V_k has been reduced to a similar voltage to the cathodes of the ECC83, allowing them to share a heater supply. We could simply insert a cathode resistor to ground, but a potential divider across the regulated output can set the required voltage and give a much lower Thévenin output resistance (15 kΩ vs 800 kΩ). The significance of this resistance is that it reduces the gain of the stage, so we want as small a resistance as possible to maintain maximum open loop gain in the regulator.

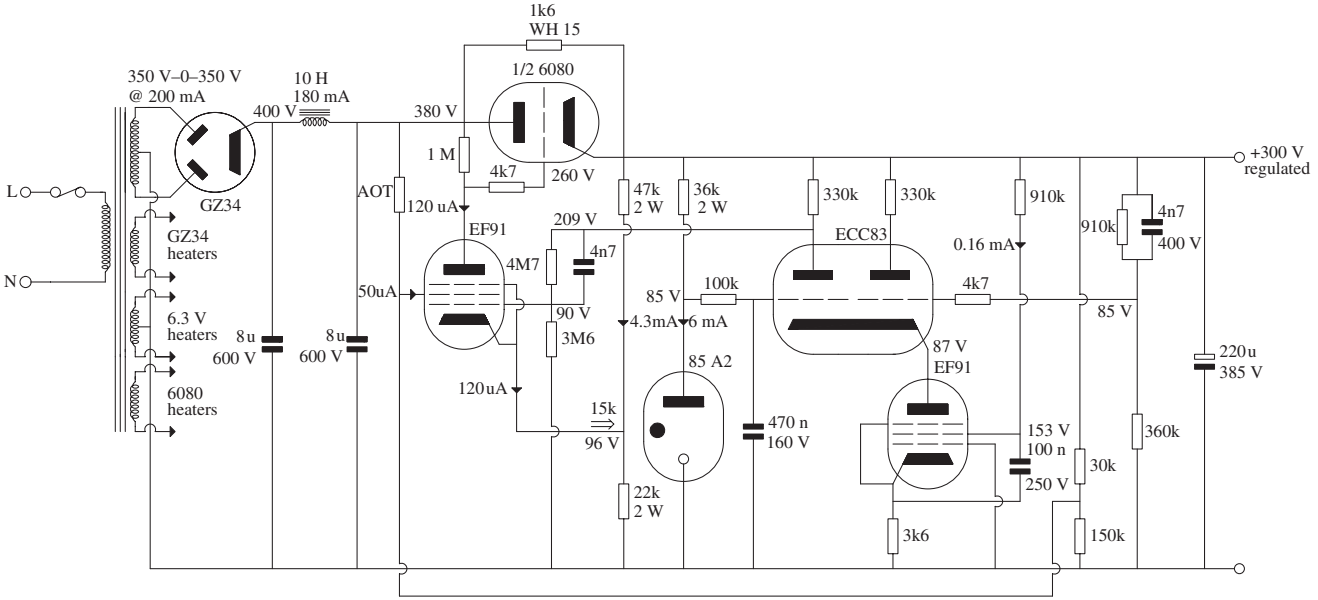


Fig. 5.38 *Optimized valve voltage regulator*

In order to couple an anode of the ECC83 to the EF91, a potential divider is needed to drop the voltage from 209 V to 90 V, thus we sacrifice ≈7 dB of DC open loop gain. However, the sacrifice is worthwhile because gain is recovered faster by dropping V_k (and reducing local feedback) on the EF91 than it is lost by the potential divider. Ultimately, the choice of V_k is usually determined by V_{hk} considerations. Nevertheless, we can recover the gain at AC by bypassing the upper resistor with a capacitor.

The ECC83 differential pair has a constant current sink tail. If we were making symmetrical split rail supplies, we would simply take a large tail resistor to the opposite supply, but a single supply needs the constant current sink.

Finally, because of the greatly increased open loop gain, the regulator has a much lower output DC resistance than before (<10 mΩ), so it needs a commensurately large bypass capacitor to maintain low output impedance at higher frequencies. A low ESR electrolytic intended for use in a switched-mode power supply is ideal.

As can be seen, much can be done to improve the basic valve voltage regulator, but the penalty is considerable complexity.

Power Supply Rejection Ratio (PSRR) to individual stages and stability

Although individual stages may be designed and interconnected to form an audio system, each stage requires HT, which will ultimately be derived from a common source. No practical source has zero output resistance, although the AC mains is quite a good approximation.

The issue of a common power supply with non-zero output resistance is crucial because it implies that as a given audio stage draws a varying supply current (in sympathy with the audio signal), a voltage will be developed across the source resistance of the supply. Although attenuated by individual stage PSRR, this voltage is now an *input* to all other stages, and if gain between stages is high (as in an RIAA stage) whilst PSRR is low, then loop gain via the power supply may be greater than unity, resulting in oscillation.

Traditional interstage filtering used a shunt capacitor to define source resistance (strictly impedance), resulting in increased source impedance at low frequencies since:

$$Z_{source} = \frac{1}{2\pi fC}$$

If the time constant of the RC network is sufficiently large, this can combine with the RC cathode bias network to turn the amplifier into a blocking

oscillator.[12] This low frequency (≈1 Hz) phenomenon was known classically as *motorboating*, but marginal stability probably went unnoticed much of the time, because loudspeakers of that time had very stiff cone suspensions and would filter it out.

Modern designs use regulators giving excellent Z_{source} down to DC, but because the error amplifier must have a response falling at 6 dB/octave in order to maintain its own stability, Z_{source} is inductive and rises with frequency, and so HF instability is a possibility.

Summarizing, any practical common supply always has a non-zero output impedance, and system stability can only be maintained if individual stages have sufficient PSRR *to that common supply*.

To aid clarity, it is useful to define two new terms:

- Intrinsic PSRR: the PSRR due to the topology of an individual stage.
- Common supply PSRR: intrinsic PSRR plus any added PSRR (by whatever means) to the common supply point.

A common cathode stage possesses intrinsic PSRR by virtue of the potential divider formed by r_a and R_L, but an E88CC operated such that $r_a = 6\,k\Omega$, and $R_L = 100\,k\Omega$ only results in an intrinsic PSRR (referred to the output) of 24 dB. Using the same valve as a μ-follower could improve this to 50 dB, and a differential pair might manage 70 dB, but used as a cascode, the 24 dB figure would be degraded to zero ($r_a \approx \infty$).

Any stage may have its common supply PSRR increased by an arbitrary amount using individual filtering, or regulation, and, apart from expense, it does not matter whether the common supply PSRR is made up mostly from intrinsic PSRR, or added PSRR via filters or regulators. Extreme methods might even include individual mains transformers and supplies for each stage, to increase common supply PSRR to the AC mains (the common supply point), whilst the use of a dedicated spur from the electricity supply company cable head would also be a means of reducing Z_{source}. The ultimate, mind-boggling, solution would be to use individual batteries for each stage.

Another method for increasing PSRR to each stage would be to use individual regulators for each stage, but even 317 regulators are expensive if we use enough of them. A cheaper method would be to design as many stages as possible to need the same HT voltage, and then isolate the stages by adding the (high) intrinsic PSRR of an op-amp to the stage's intrinsic PSRR by supplying each stage via a voltage follower op-amp. See Fig. 5.39.

The idea is that each buffer has as its input an *RC* filtered supply, and the high current gain of the buffer allows a low output resistance to drive the load,

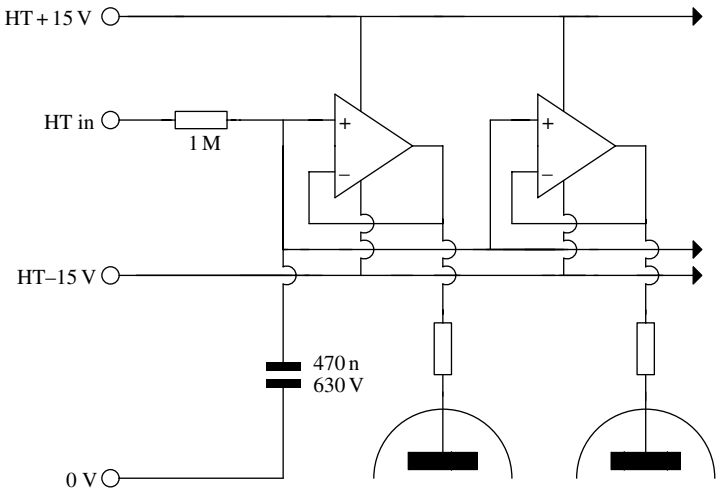


Fig. 5.39 *Use of op-amp buffers in HT supply to isolate stages*

whilst the high input resistance provides a very light load on the *RC* filter. A Junction Field Effect Transistor (JFET) would be ideal as the input buffer because of its very high input resistance. Unfortunately, JFETs tend to have a high output resistance and appalling device variation, so a bipolar emitter follower could be added to form what is occasionally known as a FETlington (FET/Darlington). This device then needs a bias network, and is rapidly becoming complicated, so why not use a JFET input op-amp as a voltage follower?

The Burr-Brown OPA2134 is a splendid JFET input, bipolar output, dual op-amp capable of sourcing up to 40 mA into the load. Its only disadvantage for our application is that it achieves its excellent specifications by consuming 4 mA per op-amp, which would be trivial if supplied from a dedicated ±15 V supply, but we are more likely to derive this current from the HT supply.

When evaluated over a 20 kHz bandwidth, the noise of a JFET input op-amp is comparable with that of an EF86 pentode. In comparison with bipolar input op-amps, JFET input op-amps tend to have a high $1/f$ corner frequency, which means that the noise rises at 6 dB/octave below ≈3 kHz. MOSFET input op-amps have an even higher $1/f$ corner frequency, and cannot be considered for audio. Although this noise will be attenuated by the anode resistance of the valve, it still contributes to the noise produced by our specially selected quiet triode.

For a power amplifier this noise is not a problem, but for a sensitive pre-amplifier it needs to be considered. The best way of dealing with it in a pre-amplifier is to make the pre-amplifier insensitive to power supply noise. This implies low r_a and high R_L to obtain maximum attenuation of HT noise, and another ploy is to use differential pairs, which further attenuate HT noise, particularly at low frequencies, which is precisely where we have our problem. These approaches will be explored in Chapter 6.

Regulator sound

Single-ended amplifiers (whether pre-amplifiers or power amplifiers) supplied from a regulator or buffer force the error amplifier to track the musical waveform. This is because the amplifier draws a current proportional to the music, and the regulator or buffer strives to maintain a constant voltage in the face of this changing current. At high frequencies, the output shunt capacitor is a short circuit and maintains a low output impedance, but at low frequencies, it is the regulator that must do the work, and cope with the (musical) current waveform. The quality of the regulator is therefore inevitably audible. Nevertheless, regulator defects are still an order of magnitude below passive supply defects.

A practical design

We now are now in a position to be able to design a complete HT and LT power supply. Since the requirements for a pre-amplifier are more critical, we will design a pre-amplifier power supply; we can then simply use whichever blocks are necessary for a less arduous application.

Specification of the power supply

Choice of HT voltage

Although the power supply should be designed to match the load, it makes good sense to consider how the power supply would have to be built to supply a given voltage, and thereby avoid specifying a power supply that would be expensive to make.

The world is full of computers. Computers are mass produced, and it is therefore economical to invest heavily in the design in order to shave

production costs. Partly for this reason, computers invariably use switched-mode power supplies. Switchers rectify the mains directly and then have a reservoir capacitor. In Europe, mains voltage is 220–240 V, resulting in $\approx 325\,V_{DC}$ when rectified. 385 V rated capacitors with low inductance are therefore readily available at reasonable cost. It makes sense to use them.

Older style HT transformers are once again becoming readily available, but isolating transformers, at any current rating we like, are available off the shelf. We now have a $230\,V_{AC}$ secondary and a $385\,V_{DC}$ electrolytic capacitor. We need a very good reason for not combining the two with a silicon bridge rectifier and using this to supply our raw HT. Allowing for the voltage dropped by the regulator and isolating buffers between stages, we could achieve a final HT voltage at each valve stage of 285 V, and this is why many of the circuits in Chapter 2 were designed to operate from 285 V.

Occasionally, 'no-compromise' designs may force a higher HT voltage, but this has expensive repercussions which will become evident at the end of this chapter.

HT capacitors and voltage ratings

If we need a higher HT voltage, perhaps 430 V for a pair of EL34, then a 450 V rated capacitor will be overstressed if mains voltage rises by 10% (as it is now allowed to do). There are two choices: we can either use a higher voltage capacitor, which will usually be a paper or plastic film capacitor and generally only available in quite low values, or we connect *equal value* electrolytic capacitors in series to obtain the required voltage rating.

Because the capacitors are connected in series, the current passing through the capacitors must be equal, so each capacitor receives an identical charge ($Q = It$). If the capacitances are equal, then their voltages must be equal ($Q = CV$).

Unfortunately, even if the capacitances are equal, the leakage currents in each individual electrolytic capacitor are unlikely to be equal, so the voltage across each capacitor will not be equal. In order to equalize the voltages, and prevent one capacitor from exceeding its rated voltage, each capacitor should be bypassed by a resistor so that the resulting potential divider chain forces the voltages to be equal. See Fig. 5.40.

The divider chain should pass at least ten times the expected leakage current of the capacitors to ensure correct operation. Typically, a 220 kΩ 2 W resistor suffices.

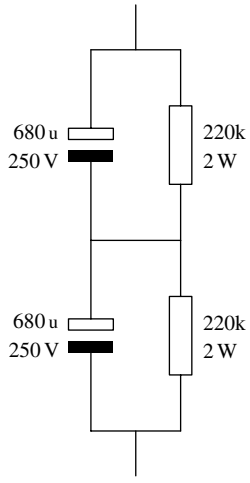


Fig. 5.40 *Bleeder resistors equalize capacitor potentials*

An even better method is to use separate HT windings and rectifier/smoothing circuits, and place the resulting floating DC outputs in series to obtain the required HT voltage. This ensures that each capacitor cannot exceed its rated voltage, but the mains transformer is now slightly more complex. See Fig. 5.41.

Can potentials and undischarged HT capacitors

Both of the previous schemes for producing a composite HT capacitor of high voltage rating resulted in one capacitor with its negative terminal stood away from ground potential. This is significant because the can potential of an electrolytic capacitor is very close to the potential of the negative terminal. Cans at an elevated voltage must not only be insulated from the chassis, but must also be properly insulated from the user to prevent shock hazard.

HT supplies represent a formidable shock hazard, and it is essential that provision is made for fully discharging the reservoir and smoothing capacitors when the equipment is switched off. The HT supply therefore needs a purely resistive discharge path to 0 V at some point, and the simplest way of providing this is to connect a 220 kΩ 2 W resistor across the reservoir electrolytic, which not only discharges the capacitor, but (provided that there is a return path) discharges subsequent HT capacitors.

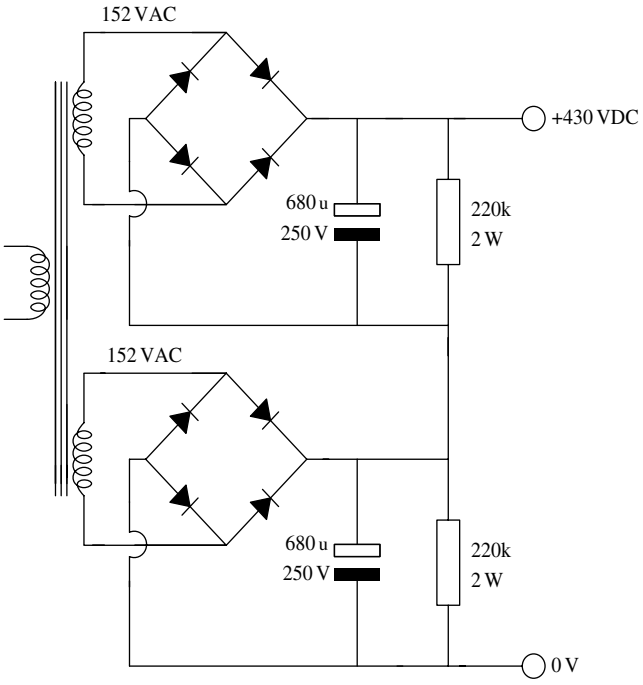


Fig. 5.41 *Achieving HT > 340 V_{DC} with electrolytic capacitors*

The switch-on surge

If we do not use a valve rectifier, the HT switches on instantly, and if this occurs whilst the signal valves are cold, it will damage their cathodes. Suddenly applying full voltage to electrolytics is not very desirable either (because of the high inrush current), so we should look to see if there is some way of avoiding these problems.

If we left the heaters permanently powered, we could apply HT instantly without causing cathode stripping. Keeping the valves warm also reduces the time taken for a pre-amplifier to reach its peak performance, as opposed to merely operating. However, keeping cathodes at their full operating temperature without drawing anode current eventually causes cathode poisoning, resulting in increased noise. The solution is to operate the heaters at $\approx$63% of their normal operating voltage in standby mode, and then apply full voltage at full switch-on.

The power supply electrolytics still need to be protected. If we suddenly apply rectified AC to the reservoir capacitor, we may be unlucky enough to

switch at the instant that the cycle is at its peak voltage. The instantaneous transition from 0 V to 325 V ($dV/dt \approx \infty$) applied to the capacitor causes a theoretically infinite current to flow, because:

$$I = C \cdot \frac{dV}{dt}$$

If, however, we always switch at the zero voltage point, then although dV/dt for a sine wave is at a maximum at this point, it is not infinite, and the inrush current is reduced.

Devices capable of performing this switching are known, predictably, as zero-voltage switching relays and are readily available. These relays require only low voltage DC to energize them, so we can remotely switch the power supply from the pre-amplifier using the permanently energized heater supply. If we are going to have a relay, why not also use it to switch mains to the rest of the equipment? The system is now becoming domestically acceptable, and requires only one switch to switch it all on.

The LT supplies

It is highly likely that we need two LT supplies, and perhaps three, because even the basic pre-amplifier has a valve with a cathode at an elevated voltage. LT supplies are not expensive to make if they are accounted for at the design stage. Adding them later is painful.

Note that *all* of the circuitry within an elevated LT supply is at least at the elevated voltage, and that it therefore represents a shock hazard if touched. Even though the circuitry only contains components rated at a low voltage, elevated supplies should be treated with as much caution as HT supplies.

RFI from external sources

RFI is Radio Frequency Interference. In its worst form it could be breakthrough from transmitters owned by taxi firms, but it could simply be a coarser treble than expected. Trying to cure RFI after it has been diagnosed is a nightmare. It is best to assume that it will be a problem, and take steps to deal with it at the design stage. Most of the precautions are constructional, and are covered in the companion book, *Building Valve Amplifiers*, but one purely electrical precaution is a mains filter.

Although the power consumed by electronic equipment may not be particularly high, the ripple current (as we found earlier) could be much

higher than the load current. A pair of large amplifiers with generous reservoir capacitors could easily draw 60 A pulses of current from the mains, even though power drawn from the mains might only be 600 W. (If 600 W seems an unlikely amount of power to be consumed by a pair of audio amplifiers, consider that the stereo 'Crystal Palace' Class A push–pull triode amplifier consumes ≈400 W from the HT supply, the output valve heaters need another ≈135 W, and ≈72 W is consumed by the remaining heater supplies and regulators.)

Most of the commercial RFI filters are rated at 16 A or less; this is not enough for audio. If we want an RFI filter, we must make it ourselves.

We now know that our basic power supply will include an RFI filter, a pair of mains transformers, a semiconductor HT rectifier/regulator, at least two regulated LT supplies, and some mains switching. We are now able to draw a preliminary block diagram of the power supply. See Fig. 5.42.

There are many choices that can be made in power supply design, usually related to whether or not you have particular components, so this design is not rigid, and is loosely designed to power the basic pre-amplifier in Chapter 7.

We need to lay down some sort of specification, otherwise we cannot design the regulators. The following values allow for extra valves to be added later, if needed, without having to rebuild the supply. You could even treat

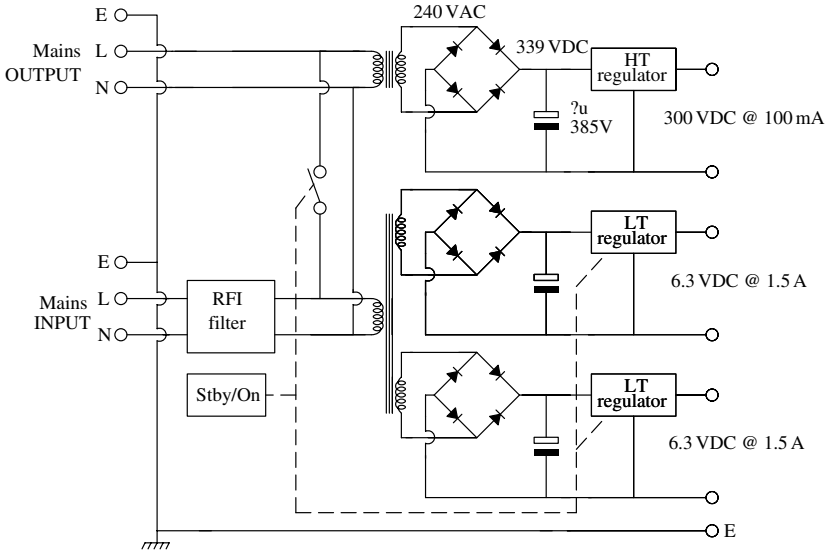


Fig. 5.42 *Preliminary block diagram of power supply*

this as a general purpose supply that will power any experimental, or permanent, pre-amplifier (this is what the author has done).

HT supply:	$300\,V_{DC}$ @ $100\,mA_{(max.)}$, $1\,mV_{pk-pk}$ ripple or less.
LT supplies:	$2 \times 6.3\,V_{DC}$ @ $1.5\,A_{(max.)}$, minimal ripple. Standby mode of 4 V

Design of the LT supplies

These are the easiest, so they make a good place to start. 1.5 A was deliberately chosen as being the maximum current that a 317 can supply. There are higher current regulators, such as the 3 A LT1085, the 5 A LM338, and the 5 A low drop-out LT1084, but high current 6.3 V linear supplies can waste as much power in the regulator as they deliver to the load. An efficient alternative is to resort to a current regulated series heater chain, which will be described later in this chapter.

Power supply design begins with regulator drop-out voltage. A 317 needs at least 2.5 V across it to avoid dropping out of regulation at 1.5 A. We therefore need 8.8 V before the regulator.

Transformers tend to have standard secondary voltages, such as $6\,V_{RMS}$ or $9\,V_{RMS}$. Assuming that we use a capacitor input filter (a choke would probably need to be custom wound), $6 \times \sqrt{2} = 8.5\,V$ which is not enough, so we must use a 9 V secondary, which will provide 12.7 V. We will use a bridge rectifier, which always has two diodes in series, so we will drop 1.4 V across the rectifier, which brings the voltage down to 11.3 V.

If we have a rectified sine wave with a peak voltage of 11.3 V leaving the rectifier, then this is the maximum voltage to which a reservoir capacitor of infinite capacitance could charge. A capacitor of finite capacitance will charge to this voltage on the peaks, but will discharge to a minimum voltage determined by its ripple voltage. The absolute minimum voltage that we can allow is 8.8 V, so the maximum ripple voltage that we can tolerate is $2.5\,V_{pk-pk}$.

Using our earlier equation that related ripple voltage to current:

$$C = \frac{It}{V}$$

The equation requires 6000 μF, so we could use 6800 μF, but this would not allow for any tolerance on capacitor value, or mains voltage variation, so 10 000 μF would be a much safer choice, resulting in 1.5 V of ripple.

A 10 000 μF low ESR capacitor is capable of delivering a very high current into a low resistance, and a wedding ring placed across the terminals of this capacitor will discharge the capacitor perfectly. The explosive arc that will be struck at the instant of discharge will vaporize some metal, resulting in flying droplets. *Do not* wear jewellery whilst working on low voltages.

1.5 V ripple is 13% of 11.3 V, and if this were the raw supply for an audio amplifier, this level of ripple would be unacceptable, but because a regulator is being used, we can allow a high ripple voltage, and benefit from the low ripple current that results.

The reduced ripple current eases rectifier current specification, and thermal considerations become more important. If the bridge rectifier drops 1.4 V, and passes an average current of 1.5 A, then it will dissipate $\approx$2.1 W (this is a rather crude approximation because $I_{\text{average}} \neq I_{\text{RMS}}$). This is a significant amount of heat to be lost from the typical 1.5 A bridge rectifier package, and they invariably become very hot. We should either use individual diodes, such as the 3 A 1N54** series, or a 4 A bridge rectifier package.

A better solution is to use Schottky diodes, perhaps the 31DQ** series, for the bridge rectifier. These have a lower forward voltage drop, reducing diode dissipation, but the main reason for their use is that they switch off cleanly, without the current overshoot exhibited by junction diodes. As mentioned early, the overshoot is an impulse that excites resonances in the transformer/rectifier/reservoir capacitor system.

The output voltage from the rectifier is 11.3 V, so V_{RRM} for each diode need only be 12 V, 50 V is commonly the lowest available rating available and allows for mains spikes, so this will be fine.

Each diode in the bridge should be bypassed with a film capacitor. 100 nF, 63 V_{DC} is a good choice, but almost anything will do, provided that the voltage rating $> V_{\text{RRM}}$ for each diode.

We should now consider how much power the 317 will be required to dissipate if the full 1.5 A load current is drawn. Ignoring ripple, the voltage applied to the regulator is 11.3 V, so the voltage across the regulator is 5 V, and the regulator therefore dissipates 7.5 W. Actually, the ripple will reduce dissipation slightly, but it is always best to be conservative in thermal calculations.

7.5 W is not an unreasonable amount of power for the (20 W) TO-220 package of a 317T to dissipate (with a suitable heatsink), so it seems to be

worth continuing with the design. (The suffix T denotes TO-220 package, whereas K is the old-fashioned TO-3 metal 'power transistor' package.)

The next step is to determine the value of resistors needed in the potential divider. Experience shows that the reference voltage tolerance on the 317 is actually very good, and that it is not usually necessary to include a variable resistor to tweak the output voltage. You may have a different view on this, but for 6.3 V, the author simply uses an upper resistor of 150 Ω, and a lower resistor of 600 Ω.

The Thévenin resistance of the 150 Ω, 600 Ω combination is 120 Ω. Theoretically, 13 μF should be used to bypass the ADJ pin to ground, but 10 μF or 15 μF will be perfectly acceptable. The manufacturers' application notes recommend that the output of the 317 be bypassed to ground with a 1 μF tantalum via a 2.7 Ω resistor.

We have now designed a fully functioning 6.3 V regulator, but need a means of putting the pre-amplifier into 'standby' mode, and reducing the heater voltage to ≈4 V to avoid cathode poisoning. The simplest way to do this is to switch a resistor in parallel with the lower leg of the divider using the normally closed contact of a relay, a 750 Ω resistor will reduce the voltage to 4.04 V.

We can now draw a heater supply circuit diagram. There will be two of these, fed from the separate 9 V windings of the LT transformer, which should preferably be a rectangular type with an electrostatic screen, but a dual chamber bobbin on a rectangular core is almost as good. It is difficult to make a toroidal electrostatic screen that doesn't leak. See Fig. 5.43.

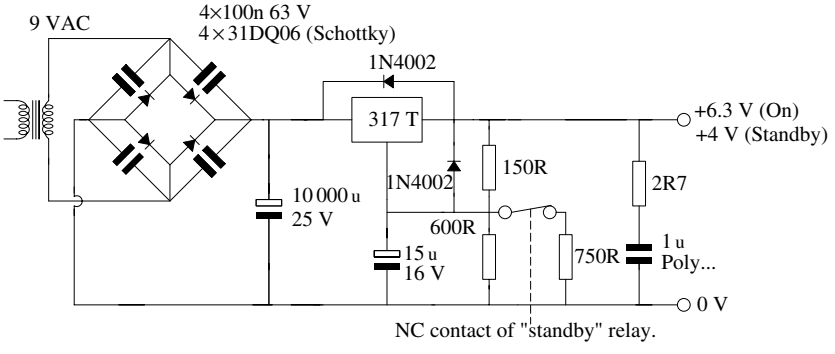


Fig. 5.43 *Practical 6.3 V regulated heater supply*

The HT rectifier/regulator

Unlike the LT supplies, the transformer voltage is already known (230 V), and so we will design the regulator to fit the raw HT voltage, rather than the other way round.

The bridge rectifier will charge the reservoir capacitor to 325 V. Although encapsulated bridge rectifiers are available that can cope with this voltage, it is safer to use individual diodes, as this allows greater spacing of the connections, with less risk of inadvertent short circuits.

If we are going to use separate diodes, it is well worth using ultra-fast soft recovery diodes such as the (1200 V_{RRM}) RHRD4120 or STTA512D, which generate less current overshoot for less time than standard junction diodes, and therefore generate less noise. Even better, silicon carbide Schottky diodes with $V_{RRM} = 600\,V$ have recently become available (SDO1060). If $V_{RRM} = 1500\,V$ is needed, but $I_{DC} < 500\,mA$, small diodes such as the BY228 that were intended for use as the *efficiency* diode (*damper* diode in the USA) in television line scan circuitry can be useful. We typically need only a very low continuous current rating (100 mA), and should select the lowest current rating above this as higher current diodes are slower and noisier.

Our regulator will be designed to produce 300 V, so we can afford to drop 25 V by way of regulator drop, reservoir ripple, and diode drop. If we use our earlier 5% criterion for ripple voltage, then the ripple voltage will be ≈17 V. However, 17 V of ripple is a much larger proportion of 25 V than we would ideally like, and it would be preferable to reduce this to 10 V or less, so a low ESR 220 μF reservoir capacitor would be ideal. It should be noted that such a capacitor, charged to 325 V, stores considerable energy, and *extreme* care should be taken when testing this circuit.

We can now consider the design of the regulator, starting with the potential divider chain. See Fig. 5.44.

If we pass 5 mA down the chain, the lower resistor must have ≈300 V across it, and a 60 kΩ resistor dissipating 1.5 W would be required. If we used a 220 kΩ 2 W resistor instead, this would only dissipate 0.4 W, which is perfectly acceptable. This has the further advantage that because the upper leg of the divider has to rise in value, the Thévenin resistance rises, and a smaller capacitor is required to bypass the ADJ pin to ground. Because the bias chain does not draw 5 mA (the minimum load current required to allow the 317 to regulate correctly), unloading the output of the supply will cause the output voltage to rise, but pre-heated valves will always present a load, so this is not a problem.

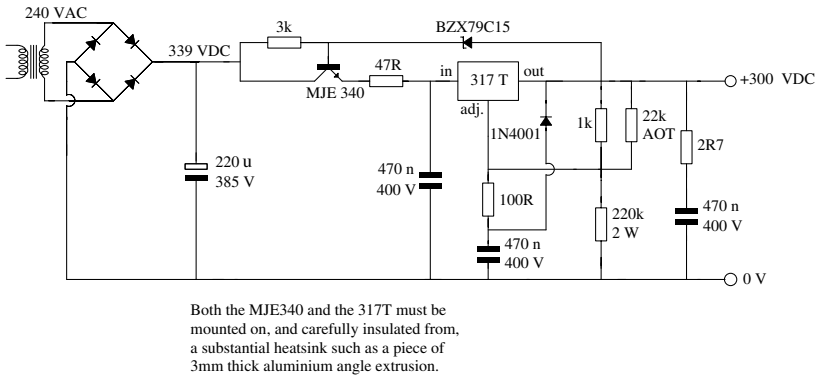


Fig. 5.44 *Practical 300V regulated HT supply*

The 220 kΩ lower resistor passes 1.358 mA, of which 50 μA is bias current from the ADJ pin of the 317. The upper resistor therefore passes 1.308 mA, and has 1.25 V across it, so a 955.7 Ω resistor is required. But the 317 reference voltage is 4% tolerance, and so there is little point in fitting this precise resistance. We could fit a variable resistor, but these are less reliable than fixed resistors, and component failure in high voltage silicon circuits is almost invariably catastrophic. The safest option is to fit a 1 kΩ resistor and leave space for it to be paralleled with an additional component that is Adjusted On Test (AOT).

Before constructing the circuit, we measure and record the exact value of the 220 kΩ 2 W resistor, which might be 221 kΩ. When the circuit is built, we measure the output voltage, which might be 290 V. By potential divider, the voltage across the 220 kΩ resistor must be 288.75 V, so 1.307 mA of current is flowing. We subtract 50 μA of bias current to find the current in the upper resistor (1.257 mA), and multiply this by the 1 kΩ resistor to find the reference voltage (1.257 V).

We can now work forwards. 298.74 V divided by 221 kΩ is 1.352 mA. We now subtract the 50 μA bias current to give 1.302 mA, and divide this into the reference voltage of 1.257 V to find the required resistor (965.6 Ω). A 27k in parallel with the existing 1k will now produce the correct 300 V HT voltage. Although this is somewhat long-winded, it is safer than adjusting a potentiometer, and the risk of subsequent failure is much lower.

The Thévenin resistance at the ADJ pin is ≈950 Ω, requiring a 1.5 μF bypass capacitor to ground. This is an expensive and bulky (400 V rated) component, so the value is commonly reduced, and 470 nF is a common choice.

In the application note that spawned this breed of regulators, a resistor was interposed between the emitter of the series transistor and the 317 to limit short circuit current. Other designs, such as those by J. J. Curcio[13] have retained this resistor for a different reason, although the value is often reduced to avoid excessive voltage drop. Adding a capacitor connected to ground at the regulator end provides RF filtering, and stability of the regulator is improved. A slight disadvantage of this approach is that there is now no possibility of the regulator surviving a short circuit to ground.

The 3 kΩ resistor in series with the 15 V Zener sets the Zener current. For low noise and maximum stability, Zeners should pass >5 mA. We know that the output of the regulator will be at 300 V, so the top of the Zener diode must be at 315 V. When we draw 100 mA from the regulator, there will be $\approx 5\,V_{pk-pk}$ of ripple on the reservoir capacitor, so the mean DC voltage is $339\,V - 2.5\,V = 336.5\,V$. Therefore the 3 kΩ resistor has $336.5\,V - 315\,V$ across it, and it passes 7.2 mA into the Zener. If you need to change the voltage going into the regulator, the value of this resistor will need to be recalculated to maintain correct Zener current.

References for elevated heater supplies

An elevated LT supply is required for any circuit with a cathode significantly above 0 V. We need an elevated supply because leakage currents develop a noise voltage across $R_{hk(hot)}$. There are two ways of minimizing the noise voltage:

- Avoid using valves with poor $R_{hk(hot)}$. Sometimes this is unavoidable, but a valve tester or purpose-made jig is required to select good examples. Because the most common cause of poor $R_{hk(hot)}$ is fluff or dust contamination during construction, it can often be burnt off by increasing heater volts by one third and monitoring $R_{hk(hot)}$ without drawing any anode current. The resistance will begin to fall, and the moment it stops changing, switch the heater off, and allow the heater to cool. With luck, when tested again, $R_{hk(hot)}$ will be significantly improved. Note that raising the heater voltage easily damages an oxide-coated cathode, but if the valve was unacceptable anyway, you have nothing to lose.
- If the DC voltage across the leaky insulation is minimized, the leakage currents will be minimized, and so will the noise.

In the following example, it has been assumed that there are two heater supplies: one for valves with cathodes at $\approx 0\,V$, and the other for valves with

cathodes at ≈130 V, so if we were to follow the RCA recommendation, we would require heaters elevated heaters by +40 V and +170 V. Although these voltages will not supply any current, they need a reasonably low AC source resistance and adequate filtering. See Fig. 5.45.

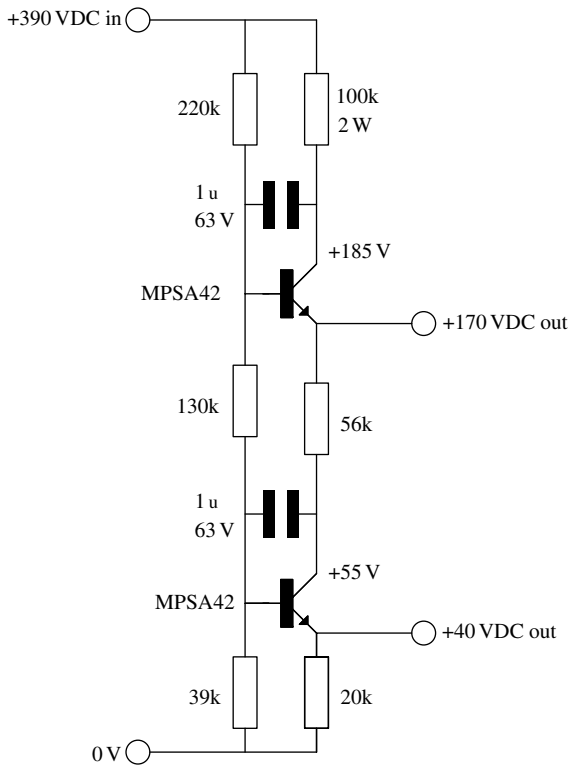


Fig. 5.45 *The THINGY, superimposing smooth DC on heater supplies*

The circuit is connected across the output of the HT supply, and is a pair of emitter followers whose output voltage is set by a tapped potential divider. The circuit is utterly non-critical of component values, and is easy to design/modify. Since we are dealing with reasonably high voltages, we will neglect V_{be} and consider the output voltages to be the same as the voltages at the tappings of the potential divider. If we neglect base current and arbitrarily set the current passing down the potential divider to 1 mA, then each resistor drops 1 V per 1 kΩ of resistance. Thus, if we want 40 V at the lower output, then 39 kΩ will be near enough for the lowest resistor. If the upper output is to be at 170 V, then the drop across the middle

resistor is $170\,V - 40\,V = 130\,V$, and a $130\,k\Omega$ resistor will do nicely. If the HT voltage is 390 V, then the upper resistor must drop $390\,V - 170\,V = 220\,V$, so a $220\,k\Omega$ resistor is required.

Although the circuit only applies a potential to the external circuit, and does not source any current, each transistor must pass some collector current, but this current is not critical, and anywhere between 1 mA and 2 mA is fine. If we set $I_c = 2$ mA, then the emitter resistor of the lowest transistor needs to be $40\,V/2\,mA = 20\,k\Omega$.

We could connect the collector of this transistor directly to the emitter of the upper transistor, but adding a collector load resistor improves the circuit's noise rejection and reduces power dissipation in the transistor. The resistor value is not in the least critical, but if we set $V_{ce} = 15\,V$ for the lower transistor, then its collector must be at $40\,V + 15\,V = 55\,V$. The emitter of the upper transistor is at 170 V, so the voltage across the collector load resistor must be $170\,V - 55\,V = 115\,V$. Since the resistor passes 2 mA, its resistance must be $115\,V/2\,mA$, and $56\,k\Omega$ is quite close enough. The advantage of including the collector load resistor is that it reduces V_{ce} (which reduces dissipation), and improves filtering.

The upper transistor also needs a collector load resistor. If we again assume $V_{ce} = 15\,V$, the collector of the upper transistor must be at $170\,V + 15\,V = 185\,V$. The HT is 390 V, so the upper collector load resistor drops $390\,V - 185\,V = 205\,V$. It passes 2 mA, so its resistance is $205\,V/2\,mA$, and a $100\,k\Omega$ resistor will do nicely. 205 V across a $100\,k\Omega$ resistor dissipates 0.42 W, so a 2 W component is required.

Filtering is achieved by placing the filter capacitor not from base to ground, which would require a high voltage component, but from base to collector. For the lowest transistor, gain to the collector $A_v = -R_C/R_E = 56\,k\Omega/20\,k\Omega = -2.8$, and Miller effect therefore multiplies this capacitor by a factor of 3.8, so that the effective value is 3.8 μF. Input resistance at the base of the transistors is approximately the Thévenin output resistance of the resistor chain and the filter cut-off frequency is therefore 1.5 Hz. The lower emitter follower sees two cascaded 1.5 Hz filters and so noise is further rejected. The value of capacitance is not in the least critical.

There is no reason to stop at just two outputs, extra output voltages can easily be derived by cascading more sections. Each section adds extra filtering, so you might choose to add a section just to improve noise rejection. Output resistance is less than $2\,k\Omega$, although supplementing each transistor with another, to form a Darlington pair, would lower this output resistance.

The author was stumped for some time in attempting to name this circuit, but eventually realized that it is a Transistorized Heater Insulation Noise Grounding Yoke (THINGY), which is what people have been calling it anyway.

Putting the blocks together

We have LT supplies and an HT supply, and we now need to put these together with some relays and transformers. We will assume that a 240 V isolating transformer was used for the HT supply, and 2 × 9 V for the LT supplies. Depending on availability of mains transformers, your supply might be different. See Fig. 5.46.

An RFI filter has been fitted at the mains input, consisting of a 130 joule metal oxide varistor, a pair of RF chokes wound on the same ferrite core, and a pair of Class X2 capacitors. X2 capacitors are the only type of capacitors that may be legally connected from live to neutral on the mains (the reason for this is that they are specifically designed to fail safely). Many RFI filters also include a pair of Class Y capacitors from live to earth, and from neutral to earth. This is not appropriate for audio equipment, as it makes the earth noisy. If electrostatic screens are available on the transformers, these should be connected directly to the point where the mains earth is bonded to the chassis.

The zero voltage switching relay simultaneously switches mains to the HT transformer and ancillary equipment. The second, conventional, relay is used to bring the LT regulators out of standby mode. It is worth using a 4-pole relay even if you only have two LT supplies, then if you add another supply at a later date, the relay contacts are ready and waiting. The 'power/standby' switch operates by grounding the lower leg of each relay coil (although the zero voltage switching relay is actually silicon, and does not actually possess a coil), this means that the raw LT voltage is not present in the umbilical cable that links the pre-amplifier to its supply and hum is not induced.

A better power supply

The previous power supply was designed to a price, whereas the following supply was simply designed to be quiet enough to supply a good RIAA stage.

The foremost priority for a pre-amplifier power supply is low noise. Not only do we want to obtain good immunity to external sources of noise (such as mains RFI), but we also want to avoid generating noise ourselves.

The worst culprit for generating noise is the rectification and smoothing process, particularly if capacitor input, so choke input supplies are almost

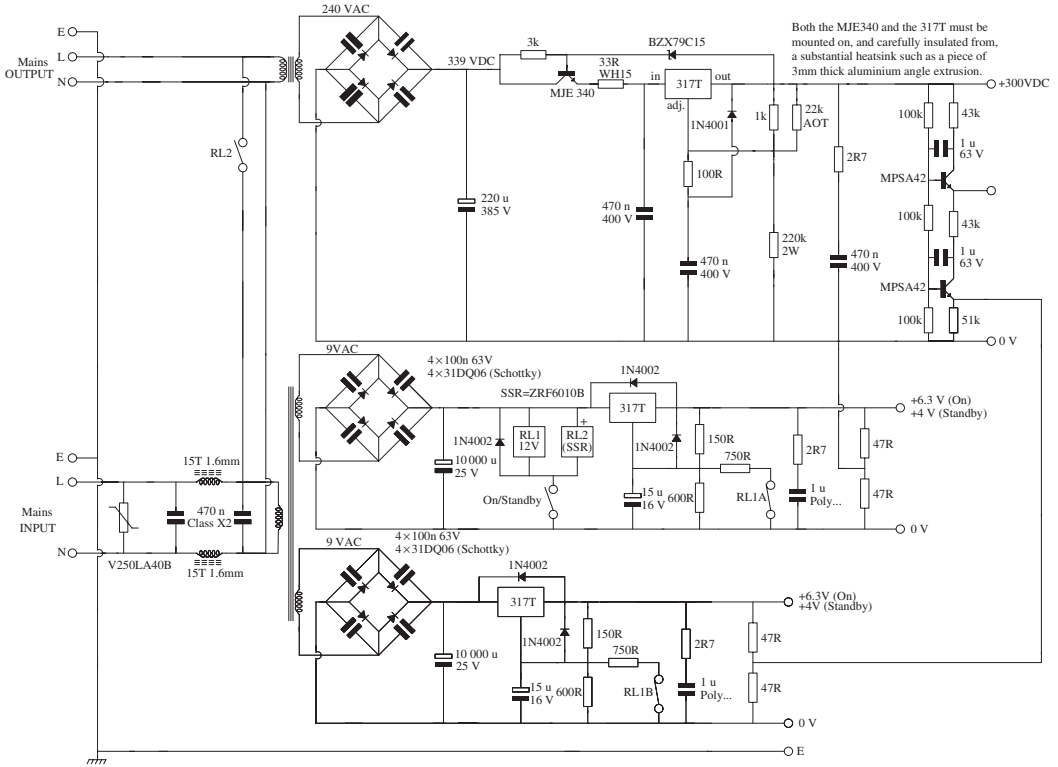


Fig. 5.46 *Comprehensive HT and LT supply with standby mode and switched mains outputs*

obligatory. Choke input HT supplies are reasonably conventional, but choke input LT supplies are unusual, and required some careful thought in conjunction with the pre-amplifier design. Power supplies should not be designed in isolation.

The LT supplies

The μ-follower RIAA stage unavoidably has an elevated LT supply in addition to the earthed LT supply, requiring two LT supplies, and uses the following lower valves:

Type	I_{heater}
EC8010	280 mA
6J5-GT	300 mA
12B4-A	300 mA

A 315 Ω resistor wired across the EC8010 heater pins matches its current to 300 mA, suggesting that a series heater chain would be possible. Current regulated series heater chains have a number of advantages compared to conventional 6.3 V_{DC} regulated supplies for parallel heaters:

- The regulator is more efficient.
- They are inherently proof against accidental short or open circuits.
- The thermal shock to the valve heaters at cold switch-on is eliminated.
- Individual heater resistances can be used as part of a staged RFI filter.
- Heater wiring resistance is now irrelevant (a complex pre-amplifier using Octal valves, perhaps consuming 6.3 V @ > 5 A would require thick heater wiring).
- Each valve heater is likely to be at a slightly higher voltage than its cathode, thus switching off the unintentional diode between emissive tungsten heater and cathode.

The disadvantages of series heater chains are:

- Any failure is catastrophic, and affects all valves in the chain. Having said that, the author has had a total of two heater failures in 30 years (one self-inflicted by flouting $V_{hk(max.)}$). Sadly, the second failure *was* in a series heater chain, and the consequential damage was horrendous.

- Thermal runaway is a theoretical possibility. As the tungsten heater warms, its resistance rises (this is true for all metals). But $P = I^2R$, so the increased resistance further increases heating. In practice, the change of resistance with temperature is quite small, and heating power is more dependent on I^2, so a current regulated supply *is* stable.

Current regulator design

As before, we also want to be able operate the heaters in standby mode, but because valve heaters do not obey Ohm's law (non-constant temperature), the required current cannot be directly calculated, even though we know that it would develop 63% of the operating voltage across the chain. A series of experiments determined that operation at 78% current is equivalent to 63% voltage for indirectly heated valves.

Series heater chains should not be connected in parallel, so we need one 300 mA constant current regulator per heater chain. The versatile 317 is ideal for the job, but instead of maintaining 1.25 V across part of a shunt potential divider chain, it now strives to maintain this voltage across a series current sensing resistor. See Fig. 5.47a.

For the regulator to operate correctly, the current sense resistor must develop 1.25 V even when it passes only the 78% standby current. Therefore:

$$R_{\text{sense}} = \frac{V}{I} = \frac{1.25\,\text{V}}{I_{\text{required}} \times 0.78} = \frac{1.25}{0.3 \times 0.78} = 5.342\ \Omega$$

Unsurprisingly, this is not a standard value, but at the expense of increasing regulator drop-out voltage, and reducing efficiency, any higher value may be used. A higher value will develop >1.25 V at the required current, which can be attenuated back to 1.25 V by a potential divider in parallel with the sense resistor, so the nearest standard value of 5.6 Ω is perfectly suitable.

In standby mode, the 5.6 Ω resistor develops 1.31 V, and in operating mode (300 mA), it develops 1.68 V. We need a potential divider chain that allows us to select between the two modes. See Fig. 5.47b.

To design the divider chain, we slightly redraw the circuit, and consider the OUT pin to be at 0 V. See Fig. 5.47c.

If the regulator is working, the ADJ pin must be at −1.25 V. If we decide to pass 1 mA through the divider chain, then the upper resistor must be 1.25 kΩ, but this is not a standard value, so we choose 1.2 kΩ, and determine our new divider chain current:

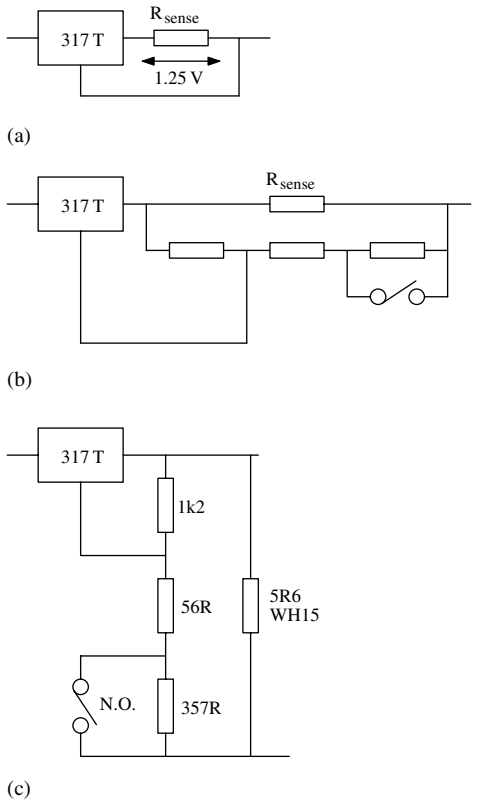


Fig. 5.47 *Using the 317 as an LT current regulator*

$$I_{\text{chain}} = \frac{V}{I} = \frac{1.25}{1.2} = 1.04167\,\text{mA}$$

The ADJ pin sources 50 μA into the chain, so the lower resistors pass 1.04167 mA + 50 μA = 1.09167 mA. The chain current contributes to load current, so the current required in the sense resistor falls to 300 mA − 1.09167 mA = 298.908 mA.

Standby mode

Standby mode is selected by short circuiting the lowest chain resistor using a relay contact. The voltage drop across the sense resistor is:

$$V_{\text{sense(standby)}} = 0.78IR = 0.78 \times 0.298908 \times 5.6 = 1.31056\,\text{V}$$

Because of regulator action, the upper chain resistor always drops 1.25 V, so the middle chain resistor must drop 1.31056 V − 1.25 V = 55.63 mV. We now have a known voltage across a resistor passing a known current, so we can find its value:

$$R_{\text{middle}} = \frac{V}{I} = \frac{0.05563}{0.00104167} = 53.4\ \Omega$$

The exact value of standby current is not critical, so we can use the nearest E24 value of 51Ω.

Operating mode

The relay contact opens and the sense resistor passes the full 0.298908 A, so it drops 1.6739 V. Between them, the two lower chain resistors must now drop 1.6739 V − 1.25 V = 0.4239 V, so we can find their total value:

$$R_{\text{middle+lower}} = \frac{V}{I} = \frac{0.4239}{0.00104167} = 406.9\,\Omega$$

56 Ω is contributed by the middle chain resistor, so the lower chain resistor = 406.9 Ω − 51 Ω = 355.9 Ω. The nearest E24 value of 360 Ω sets the heater current 0.2% high, but the E96 value of 357 Ω reduces the design error to +0.05%.

Errors and failures

In practice, by far the largest source of error is the 5.6 Ω sense resistor. In this design, the sense resistor dissipates 0.5 W. Even in free air, a 0.6 W resistor would be too hot dissipating 0.5 W, so we need a higher rated component. The obvious choice is an aluminium-clad resistor bolted to a cool heatsink, but < 5% tolerance types are difficult to obtain. A more accurate solution is to connect ten cheap 56 Ω 0.6 W 1% resistors in parallel to give a 5.6 Ω 6 W 1% component.

To eliminate design and component tolerance errors, we could make the lower resistor out of a 330 Ω fixed resistor and a 50 Ω variable resistor in

series. Not only would this allow heater current to be set precisely on test, but it would allow us to simplify design calculations by neglecting the contribution of chain current to the load. Despite this, using a variable resistor is not recommended. Variable resistors normally fail with the wiper contact open circuit, and in this circuit that would cause the regulator to turn on fully, causing the full pre-regulator voltage to be applied to the heaters, which could cause a catastrophic rise in anode current.

The standby/on relay contact has been designed to fail safely. Its contact is most likely to fail open circuit, which would hold the heaters at their normal operating temperature. Conversely, if a closed contact set full heater current, open-circuit failure would cause the cathodes to be inadequately heated at the application of HT, causing further damage.

Because of the comparative fragility of valve circuitry using series heater chains, current regulators must be designed and built to be as reliable as possible.

Specifying the LT mains transformer and choke

Working backwards from the load ($\approx$25.2 V), we drop 2 V across the sense resistor, and need 3 V across the 317, so an absolute minimum of 30 V is required before the regulator. The chosen choke had an internal resistance of 1.2 Ω, and 0.6 A (two heater chains) causes a potential drop of 0.72 V, requiring a DC component of $\approx$31 V at the output of the rectifier. The bridge rectifier adds two diode drops, although Schottky diodes will reduce this to $\approx$1 V total, so 32 V is required. Remembering that the DC component of a rectified sine wave is $0.9\,V_{in(RMS)}$, we need $36\,V_{RMS}$ from the transformer. However, this does not allow for mains voltage variation or valve heater voltage variation, so $40\,V_{RMS}$ would be a better choice.

It is easier to check a potential candidate for suitability than to specify a choke that can probably not be found. An NOS swinging choke has the following specification:

$$\mathrm{L} = 180\,\mathrm{mH}\ @\ 200\,\mathrm{mA_{DC}}$$
$$\mathrm{L} = 90\,\mathrm{mH}\ @\ 1.5\,\mathrm{A_{DC}}$$

The lower value of inductance implies the highest AC current:

$$i_{AC(\text{positive peak})} = \frac{V_{in(RMS)}}{1155L} = \frac{40}{1155 \times 0.09} = 385\,\mathrm{mA}$$

Allowing for a little current to be drawn by other parts of the LT supply:

$$i_{\text{total peak current}} = I_{DC} + i_{AC(peak)} = \approx 600\,\text{mA} + 385\,\text{mA} \approx 1\,\text{A}$$

This is less than the 1.5 A maximum rating, and the proposed choke is therefore satisfactory.

Mains transformer current specifications refer to i_{RMS} (sine wave), so we may divide our peak current by a factor of $\sqrt{2}$ to give 0.71 A, so a 30 VA transformer would be just adequate, but the author chose a 50 VA type because it was virtually the same price, and derating reduces transformer core flux density and leakage flux, which in turn reduces induced hum into nearby circuits.

The HT supply

The μ-follower RIAA stage requires 390 V @ ≈80 mA, and a 525–0–525 V 250 mA transformer and a 20 H choke salvaged from a Solartron Varipak were available. The Varipak was rated at 100 mA, so there is no need to check current rating of the choke.

The HT regulator

An early design tried a shunt HT regulator similar to that used by Allen Wright[14] for his FVP pre-amplifier, but it was destroyed as a consequence of a valve heater failure, underlining the importance of considering the entire system when designing a single block.

Shunt regulators are inherently short circuit proof, but they are vulnerable to open circuits. If the load is disconnected, the shunt element has to pass all the expected load current in addition to its normal current, and it dissipates significant power. A single heater failure in a series heater chain removes heater power from all valves in that chain, switching off *all* the valves, thus unloading the HT supply. Shunt HT regulators and series heaters are an uncomfortable combination, particularly if HT currents >20 mA are required.

Valve regulators are extremely robust, and can be made to be fairly quiet, but they invariably require multiple heater supplies, and the error amplifier ideally needs a regulated heater supply to reduce DC drift, further increasing complexity. Nevertheless, some designers swear by valve regulators – it's a matter of personal choice.

For the moment, the author still prefers the Maida regulator, although a minor improvement can be made in the potential divider chain. Previously, a 220 kΩ metal film resistor had been specified, since this value minimized wastage of HT current and power dissipation in the resistor. Now that more current is available, a 47 kΩ wirewound resistor can be used, with the advantage that it does not produce excess noise. See Fig. 5.48.

Increased V_{RRM} by multiple rectifier diodes and precautions

Because 1700 V diodes are not commonly available, the choke input HT supply uses three diodes in series to triple V_{RRM}. However, a voltage equalizing capacitor is needed across each diode to ensure that they see equal reverse voltages. The argument for this is similar to that for the voltage equalizing resistors needed for series-connected electrolytic capacitors. The diodes are in series, so they pass equal current, and therefore an equal charge ($Q = It$). The reverse voltage developed across each diode is determined by its capacitance when switched off ($Q = CV$), but this could vary, causing the reverse voltage to change. The author measured $C_{reverse} \approx 600$ pF for the STTA512D, so we swamp this possibly variable value with a 10 nF plastic film capacitor to ensure that none of the diodes exceeds its V_{RRM} rating.

When a diode is switched off, it passes a few μA of leakage current. Alternatively, we could view this as a perfect diode in parallel with a leakage resistance. Once we put diodes in series, potential divider action could cause unmatched leakage resistances to develop a voltage across a diode that exceeds its V_{RRM} rating. The problem can be avoided either by matching diodes for leakage current, or by adding a resistor in parallel with each diode that passes several times the expected leakage current. In this instance, each diode sees 589 V when switched off, so a 1 M resistor would pass 589 μA of current, which easily exceeds the diode leakage current. Unfortunately, the resistor needs to be a 2 W type to achieve the required voltage rating.

Centre-tapped transformer winding resistances

A centre-tapped transformer economizes on diodes and capacitors, but adds a new problem. Centre-tapped transformers invariably have their windings wound one on top of the other, so the average diameter of the outer winding is a little larger than the inner winding, resulting in a slightly higher copper resistance. Unless balanced by adding an external resistance to the inner winding, a ripple component at mains frequency appears at the output of the

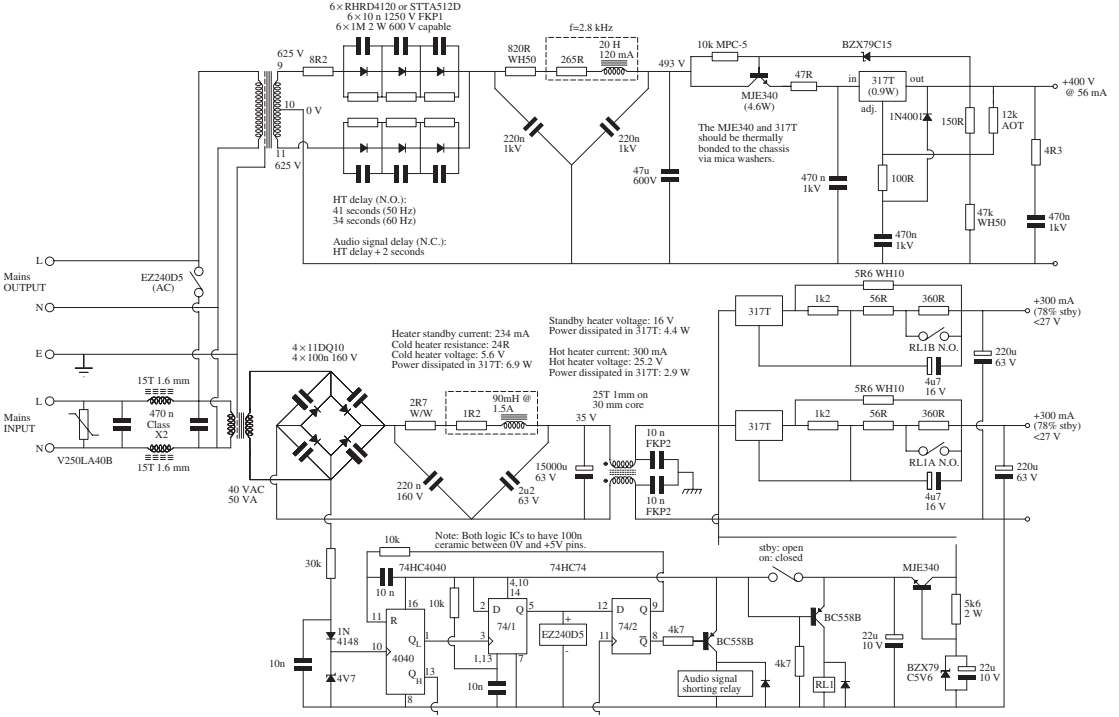


Fig. 5.48 *Complete power supply for μ-follower RIAA pre-amplifier*

rectifier, which will not be particularly well attenuated by the following smoothing. This is a very minor point, but the addition of a single cheap resistor to equalize the two resistances eliminates it.

The HT delay circuit

At the beginning of this chapter, valve rectifiers were mooted as a solution for soft starting valve electronics. Valve rectifiers are expensive, and wasteful of HT voltage, but semiconductor rectifiers normally apply the HT before we are ready for it.

As before, a zero voltage switching solid state relay is used to energize the HT transformer gently, but this relay is delayed by ≈41 seconds to allow the cathodes to warm from their standby temperature to full operating temperature.

Additionally, an output is provided for driving a relay that is normally a short circuit across the output of the associated pre-amplifier. The relay energizes ≈2 seconds after the HT appears, but releases just before the HT begins to decay, thus eliminating thumps at switch-on and off, which could damage transistor amplifiers.[15]

How the delay circuit works: we half-wave rectify one output of the LT transformer via a 30k resistor (to reduce DC flowing through the transformer) which produces 50 Hz pulses that are clipped to ≈5 V by the 4.7 V Zener. The 10 n capacitor removes RF noise that would otherwise falsely trigger the 4040 counter. Q_L of the 4040 counter changes from logic 0 (0 V) to logic 1 (5 V) after 2048 mains cycles, and the positive going edge triggers the following 74 D-type into loading the logic 1 on its D input onto its Q output, thus energizing the HT relay.

Simultaneously, the HT relay Q output is applied to the D input of the next D-type, but is not loaded to its output until Q_H of the 4040 next changes from 0 to 1, which occurs 128 mains cycles later. The inverted Q output is used to switch on the BC558B transistor and energize the audio shorting relay, or relays. The relays must be bypassed by diodes to prevent back-EMF spikes from destroying the driving transistor.

Any significant loss of power to the LT supply discharges the LT reservoir capacitor, and resets the timer circuit, thus switching off HT and muting the audio. The 'standby/on' switch applies power to the delay circuit and full heater voltage to the valves, so accidentally plugging the supply into the mains with the 'standby/on' switch switched on still delays the HT.

References

1 Watts, M. (1999) Hybrid Bridge Rectifier. *Electronics World*, August, p. 650.
2 The Edison Swan Electric Company Ltd. (1949) ESU872 Half-wave Mercury Vapour Rectifier Data Sheet. July.
3 Mullard data sheets. Gas-filled Rectifiers: General Operational Recommendations.
4 Schade, O. H. (1943) Analysis of Rectifier Operation. *Proc. I.R.E.*, Vol. 31, No. 7, July, pp. 341–361.
5 Capacitance 'Cathode Ray' (1949) Smoothing Circuits: (1) Resistance. *Wireless World*, October, pp. 389–393.
6 Cockcroft, J. D. and Walton, E. T. S. (1932) Experiments with High Velocity Positive Ions. *Proceedings of the Royal Society (A)*, Vol. 136, pp. 619–631.
7 Arthur, Kenneth (and Thompson, John) (1968) *Power Supply Circuits*, 2nd ed. Tektronix.
8 3-terminal Regulator is Adjustable. National Semiconductor Application Note 181. *Linear Applications Handbook*.
9 *1990 Linear Databook* and *1992 Linear Databook Supplement*. Linear Technology.
10 Maida, Michael. High Voltage Adjustable Power Supplies. National Semiconductor Linear Brief 47. *Linear Applications Handbook*.
11 *Type 555 Cathode Ray Oscilloscope Instruction Manual* (1961) Tektronix.
12 Kobayashi, M. (1929) US Patent No. 1,913,449.
13 Curcio, J. J. A Vacuum Tube Pre-preamp with Active Voltage Regulation. *The Audio Amateur*, 5/84, pp. 7–14.
14 Wright, Allen (1995) *The Tube Preamp CookBook*, 1st ed.
15 Self, Doug. (1996) Precision Pre-amplifier '96. *Electronics World*, September, pp. 708–716.

Further reading

Puckle, O. S. (1943) *Time Bases*. Chapman & Hall.
Reich, Herbert J. (1939) *Theory and Applications of Electron Tubes*, 1st ed. McGraw-Hill.
Tomer, Robert B. (1960) *Getting the Most out of Vacuum Tubes*. Sams. Reprinted by Audio Amateur Press (2000).

6

The power amplifier

The task of a power amplifier is to provide a fixed level of gain to a processed signal and deliver *power* into a load such as a loudspeaker. It should do this without introducing spurious signals, such as hum, noise, oscillation, or audible distortion, whilst driving a wide range of loads. Additionally, it should be tolerant of abuse, such as open or short circuits. It will be appreciated that this is not a trivial objective, and will therefore require careful design and execution if it is to be achieved.

The determining factor is the output stage. The solution adopted here dictates the topology of the remainder of the amplifier, so we will begin by investigating the output stage.

The output stage

Typical audio valves are high impedance devices and can swing hundreds of volts, but deliver only tens of milliamps of current. By contrast, a loudspeaker of typically 4–8 Ω nominal impedance requires tens of volts and amps of current. The obvious solution to this problem is to employ an *output transformer* to match the loudspeaker load to the output valve or valves.

This is where the problems start. As was hinted earlier, transformers are rather less than perfect, and the ultimate quality of a valve amplifier is limited by the quality of its output transformer. Despite this, the transformer coupled output stage is a good engineering solution, and is used in most valve amplifiers (see later for Output Transformer-Less designs).

Valves designed specifically for audio use generally have optimized configurations that are detailed in the manufacturers' data sheets. Designing output stages for audio valves from first principles is reinventing the wheel,

but an overview of the practicalities is most useful, therefore we will indulge in a brief analysis of a currently fashionable topology.

The single-ended Class A output stage

A typical transformer coupled output stage is the familiar common cathode triode amplifier using cathode bias. See Fig. 6.1.

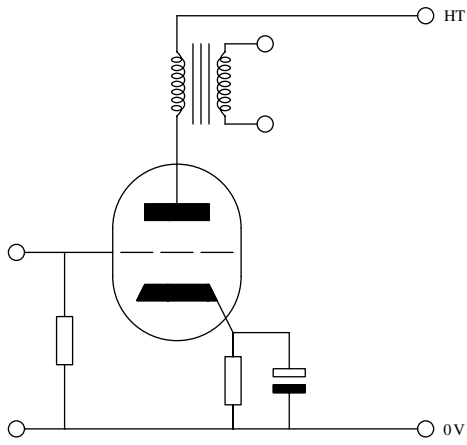


Fig. 6.1 *Single-ended transformer coupled stage*

When we investigated voltage stages, we used a loadline to choose the value of anode load, and generally optimized for linearity, rather than voltage swing; this time, we need to maximize power. For this example we will use an E182CC double triode, which might be useful as a headphone amplifier. We would normally set the operating point at the intersection between maximum continuous anode voltage ($V_a = 300\,V$) and maximum anode dissipation ($P_a = 4.5\,W$). But as there is a grid curve intersecting at $V_a = 295\,V$, the operating point has been moved for convenience. For maximum output power, the optimum load for a triode is $2 \times r_a$. In our example r_a is $3.57\,k\Omega$, so $R_L = 2 \times r_a = 7.14\,k\Omega$, and we plot this loadline. See Fig. 6.2.

$V_{gk} = -1\,V$ is our positive limit from the bias point of $V_{gk} = -13\,V$, therefore the negative limit will be $V_{gk} = -25\,V$ for a symmetric input voltage. This results in a peak-to-peak output voltage swing of $430\,V - 85\,V = 345\,V$, or $122\,V_{RMS}$, which equals 2.1 W dissipated in the load. Under these conditions, 4.5 W is dissipated in the valve, giving an efficiency of 32%.

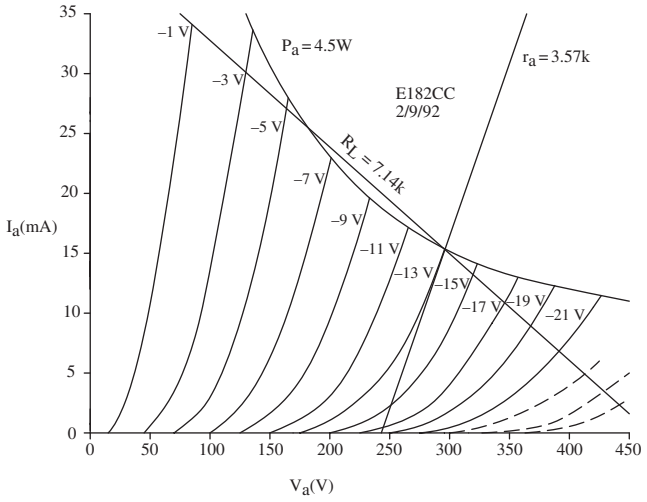


Fig. 6.2 *AC loadline for transformer coupled stage*

We should now observe some important points about the operation of this stage:

- The loadline strays into the region where $P_a \geq 4.5\,W$. Since the stage is driven only with AC (it must be, as we could not otherwise transformer couple to the load), this is not a problem. This is because although on one half cycle the anode dissipation is $\geq$4.5 W, on the other half cycle it is less, and the thermal inertia of the anode will average the dissipation out at $\leq$4.5 W.
- We set the operating point of the valve at 300 V. If the transformer is perfect, then there will be no DC voltage dropped across the primary winding, and so the HT voltage must be 300 V. Yet we have allowed V_a to rise to 430 V, which is considerably above HT. This is possible because the transformer stores energy in the magnetic flux of its core. In theory, a perfect valve could swing V_a from 0 V to 2 $\times$ HT, which is a very useful feature in a power amplifier.
- We carefully set our anode load at 7.14 kΩ, but in doing so, we assumed that the loudspeaker was a resistor. Loudspeakers are not resistive, and the transformer is not perfect, so the actual load seen by the valve will not be a precise resistance, but a complex and variable impedance.

The valve therefore sees an AC loadline that is an ellipse with its major axis roughly aligned with the theoretical resistive loadline. The gradient of the

major axis is the resistive component, and the width of the minor axis indicates the relative size of the reactive component. This means that most of the calculations we can make for an output stage are informed guesses at best, and there is little point worrying about precise values.

- Because we wish to maximize the power in the load, we have to maximize the anode voltage swing, resulting in poor linearity. We could improve linearity by increasing the value of the anode load and plotting another loadline, but this will reduce available output power.

Although the linearity of single-ended stages is not good, the distortion produced is mostly 2nd harmonic, which, as we observed earlier, is relatively benign to the ear. We can estimate the percentage of 2nd harmonic distortion from the following formula:

$$\%D_{2\text{nd harmonic}} \approx \frac{V_{\text{quiescent}} - \frac{V_{\text{max.}} + V_{\text{min.}}}{2}}{V_{\text{max.}} - V_{\text{min.}}} \times 100\%$$

In our example, $V_{\text{max.}} = 430\,\text{V}$, $V_{\text{min.}} = 85\,\text{V}$, and $V_{\text{quiescent}} = 295\,\text{V}$, resulting in 11% 2nd harmonic distortion at full output. Clearly, >10% distortion is not Hi-Fi, but the attraction of single-ended amplifiers is that their distortion is always directly proportional to level, and so at one tenth output power, the distortion would be ≈1%, and so on. Since, most of the time music requires very little power, it is often argued (oddly enough, by single-ended enthusiasts) that it is the quality of the first watt that is important, not the remainder that are rarely used. The distortion could be reduced by negative feedback, but this technique is almost universally shunned by the supporters of single-ended amplifiers, so single-ended amplifiers not only produce high distortion, but also tend to have an output resistance of half the assumed load resistance.

The significance of high output resistance

The vast majority of modern loudspeakers use moving coil drivers in sealed or reflex boxes. The theory of the interaction between moving coil loudspeakers and their enclosures was introduced in a landmark series of papers by A. N. Thiele and R. Small in the *Journal of the Audio Engineering Society* in the early 1970s. A closed box is a 2nd order high-pass filter, whereas a reflex box is 4th order, although it can be designed to look like 3rd order. The crucial point is that Thiele and Small showed that the Q of the high-pass filter could

be precisely set by the series contribution of voice coil resistance, crossover resistance, and amplifier output resistance – which is normally assumed to be zero. Typical single-ended amplifiers make a mockery of the zero output resistance assumption and cause the loudspeaker to produce a peaked bass response that the loudspeaker designer did not intend.

Reflex loudspeakers developed prior to Thiele and Small relied on the mechanical damping contributed by the suspension of the loudspeaker to determine bass response and made few assumptions about amplifier output resistance. Horns rely on the transformed air load for damping, so they too are tolerant of high amplifier output resistance. As a further bonus, both of these types of loudspeakers are sensitive, so they are popular with afficionados of single-ended amplifiers.

Bass is produced by moving a large volume of air, and requires a large, rigid, and consequently heavy, cone. Treble is produced by accelerating and decelerating the cone or dome many times a second, requiring low mass. The requirements for bass and treble reproduction are contradictory, so most loudspeaker designers prefer to use optimised drivers for specific frequency bands supplied from an electrical filter known as a crossover. Nevertheless, some designers feel that the practice of multiple drivers and crossovers is so fundamentally flawed that they attempt the design of full-range drivers. In practice, the treble response tends to be a little peaky after 10 kHz, and the low frequency resonance limits bass to ≈100 Hz, but this is a sizeable proportion of the audio band. Fortuitously, the motion of a full-range driver's low mass cone is easily damped, and when mounted on an open baffle, the mechanical damping of the suspension is frequently sufficient. Thus, a full-range driver mounted on an open baffle can be an ideal match for a single-ended amplifier.

Transformer imperfections

When we plotted the single-ended loadline, we treated the output transformer as being perfect, but we should now consider how its imperfections will affect the stage. Unfortunately, we pass a constant magnetizing current ($I_{quiescent}$) through the primary of the output transformer. In order for the core not to saturate, which would cause odd harmonic distortion, we need a large gapped core. Another method of avoiding core saturation is to reduce the number of primary windings, which reduces the magnetizing effect of the quiescent current, but also reduces primary inductance.

Usually both methods have to be used, which results in a physically large transformer of low primary inductance at the operating point. Because the

transformer is so large, it has correspondingly large stray capacitances, and the high frequency performance is also compromised. Typically, the transformers used in this way are large, expensive, and have compromised bandwidth in comparison with a push–pull transformer.

It might therefore be thought that the single-ended transformer coupled power stage is a complete non-starter, but, curiously, this is not so. If we look at the hysteresis curve for transformer iron, this offers a clue to this topology's recent resurgence in popularity. See Fig. 6.3.

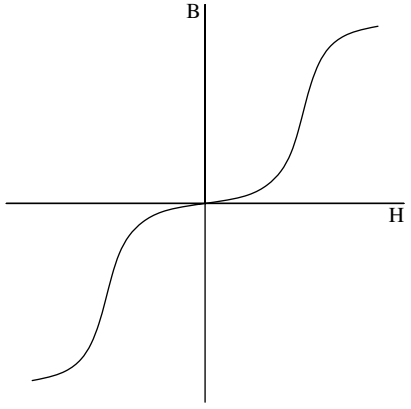


Fig. 6.3 *Exaggerated BH curve of iron*

When used as a transformer, the hysteresis curve may be treated as a transfer characteristic showing the relationship between V_{in} and V_{out}. If there were no DC current flowing through the transformer, then an AC signal would swing symmetrically about the origin. If we look at the small-signal performance, we see that there is a kink in the characteristic around the origin where the slope of the curve is reduced. Since the core's permeability is proportional to the slope, the transformer has reduced primary inductance (L_p) at low levels. At low frequencies, reduced L_p reduces gain and increases distortion in the output stage.

The cause of the kink is that the individual magnetic domains that make up the core have stiction in reversing the polarity of their magnetism. (The same effect is true in electrostatics, in that there is stiction in reversing electrostatic charges in polar materials such as polyester and polycarbonate.) A recently popular solution known as *pinstriping* uses a mix of steel and mumetal laminations to make up the core. Mumetal has a much higher initial permeability, so it maintains high L_p at low levels, but saturates quite quickly, at which point the

steel takes over, so pinstriping can improve the initial permeability of the core. Unfortunately, mumetal is fragile and more expensive than steel.

Alternatively, by passing the valve's quiescent current through the transformer, we sidestep the problem of low initial permeability and the transfer characteristic is more linear, which perhaps explains the claims for excellent midrange detail in this breed of amplifiers.

Although the transformer has a low primary inductance, suggesting a poor bass response, a well-designed core is less likely to saturate at low frequencies, since it had to be oversize and gapped to accommodate the quiescent current. Because of this, L_p is nearly constant from full AC output power to zero AC output. Provided that the loudspeaker is carefully matched, a good subjective bass quality can be achieved, because it does not change with level.

Unfortunately, we can make no excuses for the HF performance. The large, leaky output transformer produces significant losses at HF, although excellent construction helps matters.

In order to reduce distortion, single-ended amplifiers typically use true triodes with directly heated cathodes, such as 2A3, 300B, 211 and 845, rather than beam tetrodes or pentodes connected as triodes. Unfortunately, directly heated cathodes are prone to hum when fed by AC from the mains transformer heater winding.

To sum up: single-ended triode amplifiers have good low-level performance, the midrange is good, but they require careful loudspeaker matching to make the most of their bass performance and low power (usually <10 W), and they are outrageously expensive due to the cost of the output transformer and the esoteric valves used. They are also surrounded by hyperbole.

The single-ended pentode amplifier found in a 1960s Dansette record player shares none of these characteristics except low output power.

Classes of amplifiers

The 'class' of an amplifier refers to the proportion of quiescent anode current to signal current. Until now, we have only looked at Class A amplifiers, although the fact was not explicitly stated. If we relax that restriction, we will need some definitions.

Class A

Anode current is set at a level such that, even with maximum allowable input signal, anode current never falls to zero. In other words, the valve

never switches off. (Maximum theoretical efficiency for sine wave output is 50%.)

Class B

There is zero quiescent anode current, and anode current only flows during the positive half cycle of the input waveform. The valve is therefore switched *off* for the negative half cycle of the input waveform, and considerable distortion of the signal occurs, since it has been half-wave rectified. Additional measures will need to be taken to deal with this problem. (Maximum theoretical efficiency for sine wave output is 78.5% for a push–pull Class B amplifier.)

Class C

Anode current flows for *less* than half a cycle of the input waveform. This method is only used in radio frequency amplifiers where resonant techniques can be used to restore the missing portion of the signal, and results in even greater efficiency and distortion than Class B.

Radio frequency engineers refer to the *conduction angle* to specify the proportion of time in which anode current flows. Using this description, we see that Class A amplifiers have a conduction angle of 360°, Class B is 180° and Class C < 180°. The transition between Class A and pure Class B is quite broad, and so there is an intermediate class known as Class AB. See Fig. 6.4.

In Fig. 6.4, the transfer characteristic of the output device is assumed to be perfect, so the input sine wave is simply reflected through the diagonal transfer characteristic to produce the output. In Class B operation, the bias voltage cuts off the valve, and it is only on positive half cycles that the signal is able to switch the valve on. It will be noticed that the output waveform of the Class B stage is very similar to the power supply waveforms in Chapter 5, and for the same reason, half-wave rectification is taking place.

Note that as bias voltage is increased negatively, the conduction angle falls.

In audio, we normally refer to currents rather than conduction angles, and there are subdivisions of classes defined by the *grid* current of the valve. (RF engineers are unable to do this since they invariably operate with grid current to maximize efficiency at the expense of linearity.)

Class *1

Grid current is not allowed to flow. Many of the larger ($\geq$50 W) classic amplifiers were push–pull Class AB1.

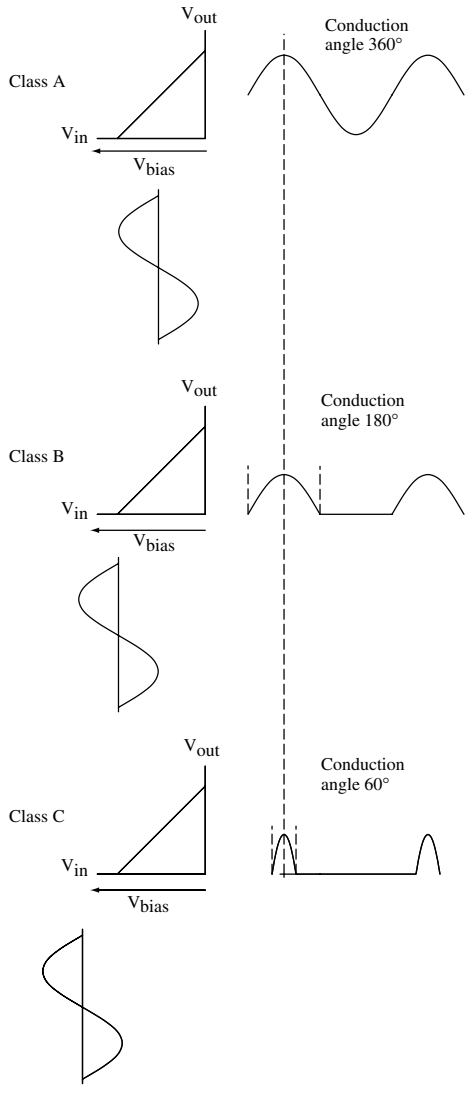


Fig. 6.4 *Relationship between anode current and input signal for Class A, B, C*

Class *2

The input signal is allowed to drive the grid positive with respect to the cathode, causing grid current to flow. This improves efficiency, since the anode voltage can now more closely approach zero, which is particularly

relevant to triodes. At the onset of grid current, the input resistance of the output stage falls drastically, (possibly approaching 1/*gm*), and the driver stage needs a very low output resistance if it is to maintain an undistorted signal into this extremely non-linear load without distortion. One way of alleviating this sudden onset of distortion is to use such a low value of grid leak resistor that it draws more signal current than the expected grid current, thus when grid current flows, the proportionate change in load current (and therefore distortion) is small. Some modern single-ended amplifiers claim to operate in Class A2, whilst transmitters are commonly Class B2.

It will be apparent that efficiency can only be improved at the expense of linearity.

The push–pull output stage and the output transformer

We saw that the Class B stage introduced considerable distortion by half-wave rectifying the input signal. Clearly, this is a disadvantage for a Hi-Fi amplifier, since we require linearity.

Suppose, however, that we had two Class B valves, one fed directly with the input signal, and the other with an *inverted* signal. During time t_1 the upper valve conducts, whilst the lower is cut off, and during t_2, the situation is reversed. See Fig. 6.5.

So far all that we have achieved is to ensure that any *one* valve is switched on, no matter what the incoming phase. By inverting one output and

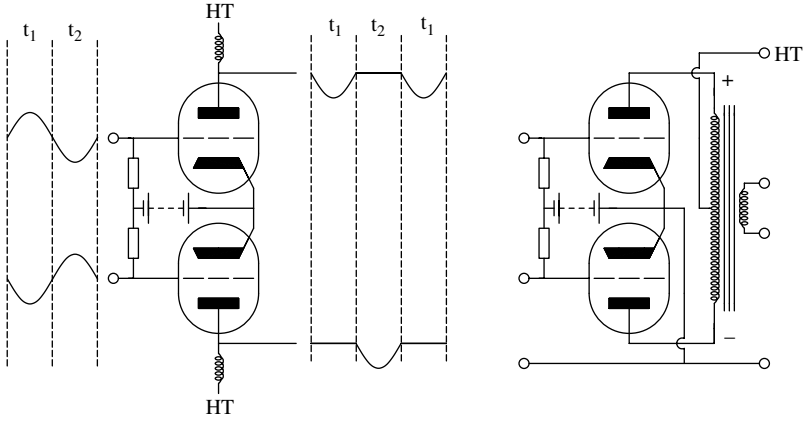


Fig. 6.5 *Summation of Class B signals in output transformer*

summing it with the other in the output transformer, we can recreate the shape of the original input waveform. The inversion is performed by reversing the connection of one winding, and is marked on the diagram with + and − symbols.

Whether achieved by a transformer or by a direct coupled series amplifier, such as a White cathode follower, this form of connection is known as *push–pull*, and is the only way of approximating linearity in a Class B amplifier.

Unsurprisingly, this dissection of the signal and its subsequent restitching is rather less than perfect, and pure Class B is rarely used because of the distortion generated at the crossover region, where one valve takes over from the other. In practice, some quiescent current is allowed to flow in an attempt to smooth the transition, resulting in Class AB operation. The theoretical optimum bias voltage for a Class AB amplifier is found by extending the linear part of the transfer characteristic until it intersects the V_{gk} axis. However, practical valves do not operate linearly down to cut-off and then suddenly switch off, so individual differences between valves mean that the ideal point is ill-defined, and crossover distortion is not eliminated.

Push–pull output stages can also be used for Class A amplifiers, giving additional advantages.

Because of the reversal of one winding, the magnetizing flux caused by the quiescent anode currents cancels (provided that they are equal). Because the transformer core only has to handle signal current, it can be far smaller for a given power – this is the *main* reason for using a push–pull output stage in a Class A amplifier.

Since the core is small, it is important that the quiescent anode current of each valve *is* identical, otherwise DC magnetization of the core will generate odd harmonic distortion. This can be done by having an adjustment for DC balance in the bias circuit, or by using pairs of valves with matched anode currents. See Fig. 6.6.

P2 sets total anode current, whilst P1 adjusts DC balance by biasing one grid more or less positively than the other.

If the core should become permanently magnetized (perhaps by failure of a valve), it will need to be degaussed, or it will generate additional (and unnecessary) distortion. This can be done by applying a sufficiently large alternating magnetic field to the core to saturate it both positively and negatively, and then reducing the field to zero over a period of about 10 seconds. In practice, the low remanence of typical transformer core materials means that this procedure is unlikely to be necessary.

A useful consequence of the reduced size of transformer is an improved HF response due to the reduction of stray capacitances.

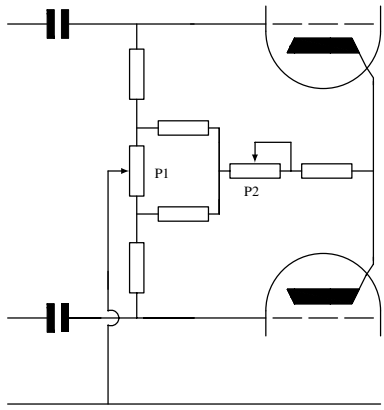


Fig. 6.6 *DC balance adjustment*

Not only does quiescent anode current cancel in the transformer, but power supply hum also cancels, since it is in phase in each winding. Improved tolerance to power supply hum allows a cheaper power supply.

Additionally, even harmonic distortion, caused by unequal gain on positive and negative half cycles, is cancelled, whilst odd harmonic distortion is summed. Since triodes generate primarily even harmonic distortion, this is useful, but pentodes generate primarily odd harmonic, and will therefore require considerable (>20 dB) negative feedback to reduce the distortion to acceptable levels. Cancellation of even harmonic distortion will only be achieved if both windings are fed identical signal levels by their valves, so some amplifiers have adjustments for AC balance, whereas others specify pairs of valves matched for gain. See Fig. 6.7.

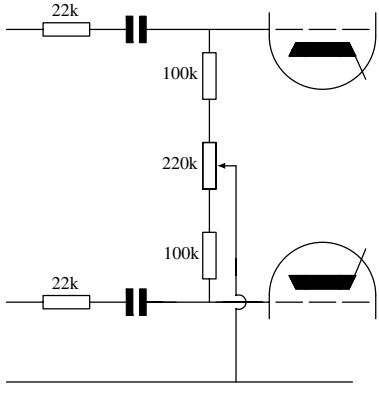


Fig. 6.7 *AC balance adjustment*

Modifying the connection of the output transformer

We have mainly considered triodes, and given pentodes scant regard because they produce copious amounts of odd harmonic distortion. But if we imagine the output transformer primary as a set of windings that could be tapped at any point, we see that for pentode operation g_2 would be connected to the centre tap (0%); whereas for a triode, it would be connected at the anode (100%). See Fig. 6.8.

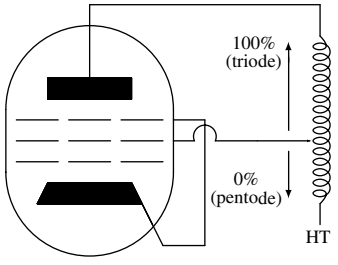


Fig. 6.8 *Blumlein or 'ultralinear' output stage*

What would happen if we were to tap at an intermediate point? This question was asked in 1951 by David Hafler and Herbert I. Keroes,[1] and an amplifier named *ultralinear* became synonymous with an output stage that had been invented by Alan Blumlein[2] in 1937. The popular tapping points are 43% (minimum distortion), and 20% (maximum power). This method became almost universal in the final days of valve supremacy, since it combined the efficiency and ease of driving the pentode with much of the improved linearity of the triode. It should be noted that:

$$I_a \propto V_{g2}^{\frac{3}{2}}$$

And as a consequence, negative feedback at g_2 is not as linear a process as one might wish. Nevertheless, almost all power amplifiers using pentodes in the output stage use this scheme because it is far superior to pure pentodes.

Up until now we have placed the transformer in the anode circuit, but we could place the same transformer in the cathode circuit to form a cathode follower, resulting in an extremely low output resistance from the valve. As an example, a pair of EL34, connected as triodes, would each have an anode resistance of about 900 Ω, but used as cathode followers, the driving resistance would be a tenth of this at 90 Ω. Reflected through the transformer, the

output resistance seen by a loudspeaker would be a fraction of an ohm even before global negative feedback.

Unfortunately, there are two crippling disadvantages to this topology. First, although the output stage is excellent, we have transferred its problems to the driver stage. Each output valve now swings $\approx 150\,V_{RMS}$ on its cathode, and has a gain <1, requiring $\approx 500\,V_{pk-pk}$ to drive it! This can be done, but it is not a trivial exercise to design the driver stage, since we must either use transformer interstage coupling, or a resistive anode load requiring a high HT voltage. The commercial solution to the high HT requirement is to bootstrap the driver from the output transformer, but this technique easily compromises HF stability.

Second, the high voltage on the cathode of the output valves severely strains the heater/cathode insulation, and can cause premature heater failure. Connecting the heater to the cathode solves this problem, but requires individual heater windings for each half of the output stage (to avoid shorting k_1 to k_2), and forces each valve to drive the interwinding capacitance (≈ 1 nF) of the mains transformer. A high frequency heater supply via a small transformer with individual windings would solve the problem, at the expense of possible RFI and increased complexity.

Alternatively, there are some valves whose heater/cathode insulation can withstand 300 V, such as the 6080/6AS7G. Because this valve has such a low r_a, its optimum load resistance is fairly low, and its output voltage at full power is also quite low, reducing the strain on heater/cathode insulation. Unfortunately, μ is also very low, so the gain of the output stage is substantially less than unity, and the driver stage has to be quite special. See Fig. 6.9.

As can be seen, a complex power supply would be required, simply to produce 6 W. Admittedly, the driver could cope with a number of 6080 in parallel, but the cure still seems worse than the complaint. The only reason that this design survived to the drawing stage is that the output stage should be quite tolerant of poor output transformers; conversely, a good output transformer would allow very good performance. The amplifier uses triodes throughout, whose distortion is predominately 2nd harmonic, but this is cancelled by push–pull action, so the amplifier relies on accurate balance rather than global feedback to reduce distortion, allowing a balanced input.

Another possibility is to use a distributed load, where part of the load is in the cathode, and part in the anode. This was used successfully by Quad in the UK and McIntosh in the USA. This relaxes the driving requirements whilst retaining some of the benefits of local feedback. See Fig. 6.10.

An interesting version of the distributed load technique was used in the Michaelson & Austin TVA10 which had equal anode and cathode loads.

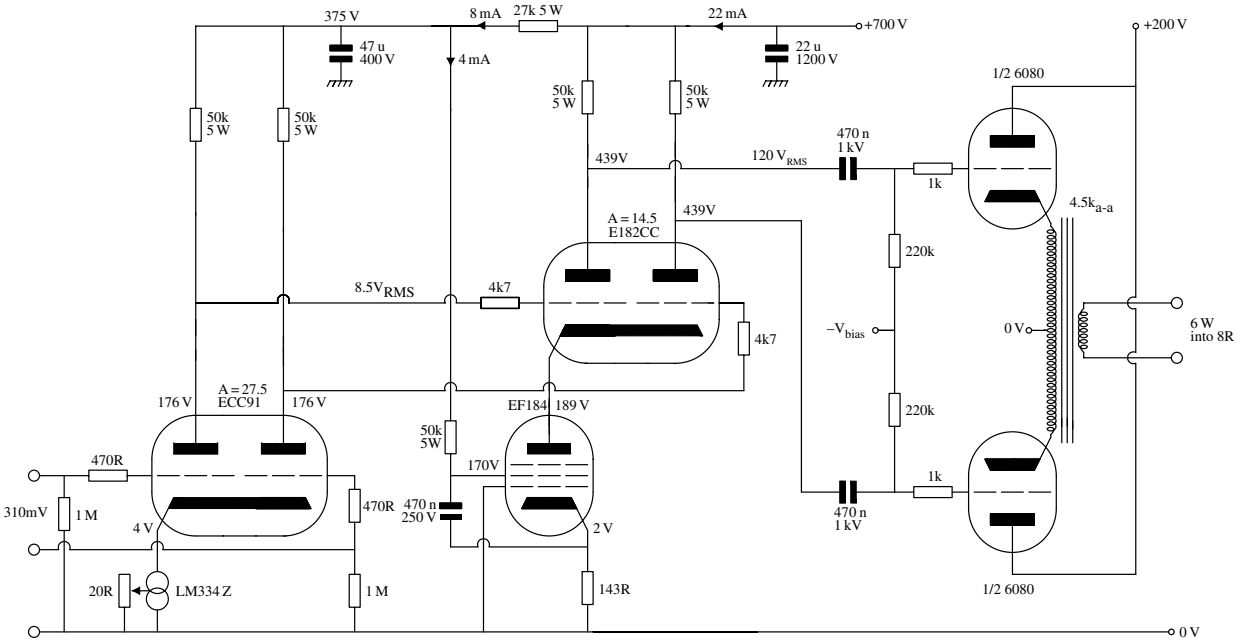


Fig. 6.9 *Amplifier with cathode follower output stage*

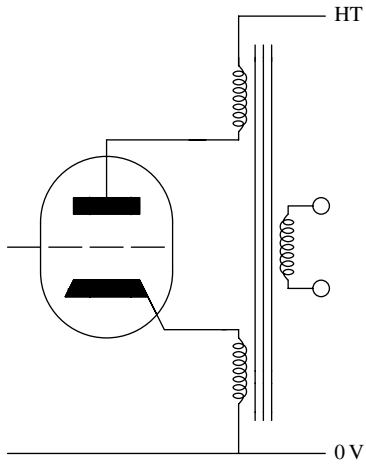


Fig. 6.10 *Quad II output stage (aka McIntosh configuration)*

Because the anode and cathode windings for each valve are wound in anti-phase (the dot signifies the start of the winding), the quiescent anode current cancels, and there is no need to match valves for anode current. Sadly, the drive requirements are almost as severe as for the cathode follower, and this technique has not been widely used. See Fig. 6.11.

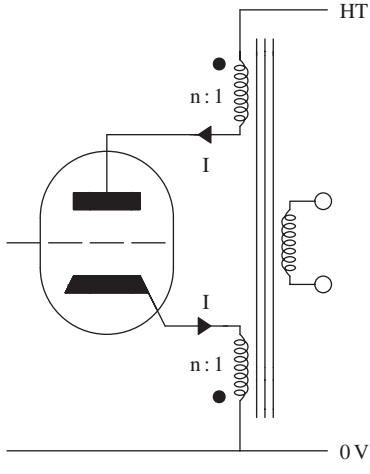


Fig. 6.11 *Complete cancellation of DC current in output transformer using only one valve*

Output Transformer-Less (OTL) amplifiers

Almost all of the different output stage configurations were devised in an effort to reduce the adverse effect of the output transformer, so it is not surprising to find that there have been some designs that dispense with the output transformer. These are often known as Futterman[3] amplifiers (who patented the notion), or OTLs.

Driving low impedance loads directly is not natural for a valve, so radical approaches are needed. Special valves have to be used, which invariably were not designed for audio, and they therefore have extremely questionable linearity and consistency. Examples are the 6080/6AS7G double triode, series regulator valve, and television line scan output valves such as the PL504 and PL519 pentodes. Efficiency is generally on the low side of appalling. Output stages invariably use paralleled White cathode followers with plenty of global feedback to reduce the output resistance. See Fig. 6.12.

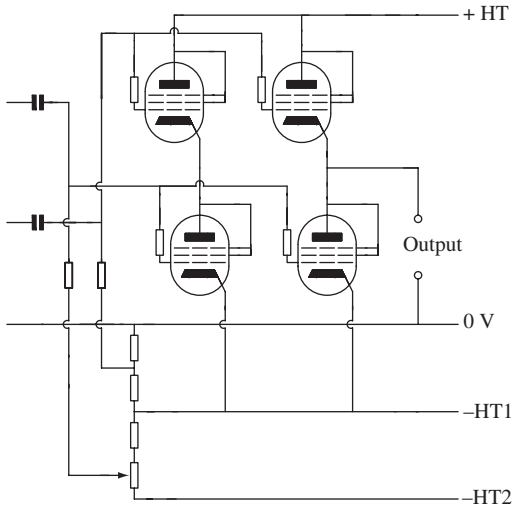
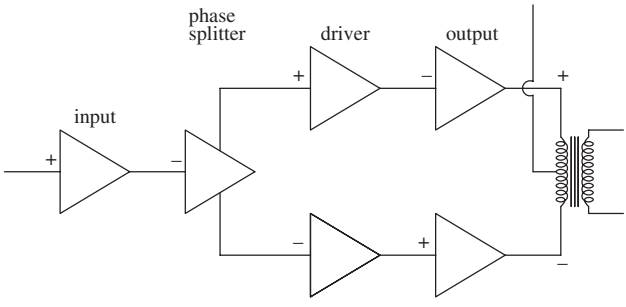


Fig. 6.12 *OTL output stage (paralleled White cathode followers)*

These amplifiers are quirky in the extreme, yet some designers think that the problems of output transformers are so severe that they persist in making successful OTL amplifiers.

The entire amplifier

Having looked at the problems of the output stage, we can now consider the support circuitry in detail. The output stage is insufficiently sensitive to be driven directly from a pre-amplifier, so it needs additional gain. If it is push–pull, it will need a phase splitter. Since linearity is unlikely to be ideal, we will probably need global negative feedback, which will further reduce gain, and this will need to be restored. A complete system might therefore comprise an input stage, a phase splitter, a driver stage, and the output stage. See Fig. 6.13.



Note : +, – indicates absolute polarity at that point.

Fig. 6.13 *Block diagram of complete power amplifier*

Although a Class A output stage is a constant resistive load, a Class AB2 output stage heavily loads the driver stage when drawing grid current, so its driver stage would need very low output resistance and be able to source significant current to drive this load without distortion.

Unlike the output stage, and possibly the driver stage, the remainder of the stages in a power amplifier will be loaded by predictable resistive loads. It is therefore possible, and desirable, to design these stages with great care in order that they should not degrade the performance of the entire amplifier. We will mainly consider push–pull amplifiers, since they make up the majority of designs, although their design principles may perfectly well be applied to single-ended amplifiers.

The driver stage

Unless the amplifier is quite low power, it will require a dedicated driver stage. We need a stage with good linearity, low output resistance, and good output voltage swing.

The differential triode pair is the ideal choice. The output stage probably requires about 25 V_{RMS} to each grid, and has an input capacitance of 40 pF, or more. An output resistance of 10 kΩ coupled to an input capacitance of 40 pF gives a high frequency cut-off of 400 kHz, which is perfectly acceptable. Since $r_a \approx R_{out}$, the high μ valves, which tend to have a high r_a, are probably unsuitable.

In a properly designed power amplifier, the output stage should be the limiting factor, so we ought to design the driver stage to have at least 6 dB of overload margin. This requirement probably rules out our favourite valve, the E88CC. Very few commonly available valves satisfy our requirements.

Dual triodes suitable as drivers:

Type	r_a	Comments
*SN7/*N7	≤10 kΩ	Lowest distortion
ECC82	≤10 kΩ	13 dB worse distortion than *SN7/*N7
E182CC	≤5 kΩ	Good on paper, but can sound strident
6BL7	≤3 kΩ	Cripplingly high C_{ag}
6BX7	≤2 kΩ	Capable of driving 845. Robust

The *SN7/*N7 family is by far the most linear of the previous selection, and if the *SN7GTA or *SN7GTB versions are used, $V_{a(max.)} = 450$ V. The 6BX7 and 6BL7 were designed for use as the field scan amplifier in televisions, but field scan amplifiers are required to be non-linear, so audio performance is variable. In a distortion test of thirty 6BX7 triode sections, the author found that distortion varied by a factor of four between samples and that very few envelopes contained a pair of low distortion triodes.

Single pentodes suitable as drivers (when triode connected):

Type	r_a	Comments
EF184	≤5 kΩ	Cheap and really plentiful. μ = 60
N78	≤3 kΩ	Uncommon, but cheap
A2134, EL84	≤2 kΩ	NOS EL84 extinct, but current Chinese production fine

Even lower driving resistance can be provided by an additional direct coupled cathode follower stage, which also has the advantage of buffering the differential pair from grid current effects. See Fig. 6.14.

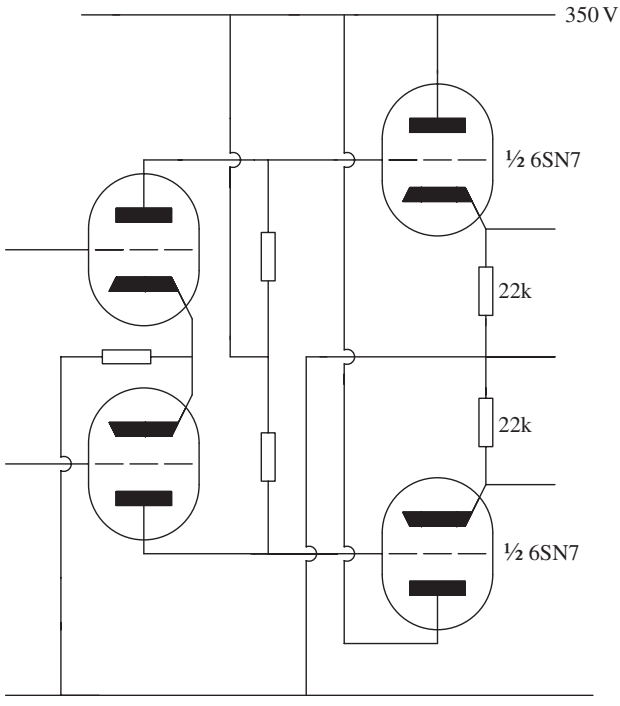


Fig. 6.14 *Driver stage using differential pair direct coupled to cathode followers*

The phase splitter

The driver stage is always preceded by the phase splitter, and traditionally the two stages have been combined – although as we shall see, this is not always a good idea. Design of the phase splitter is crucial to the success of a push–pull amplifier, so we will look at this in detail.

The phase splitter converts a single-ended signal into two signals of equal amplitude, but with one of inverted polarity. There are three fundamental ways of achieving this goal. See Fig. 6.15.

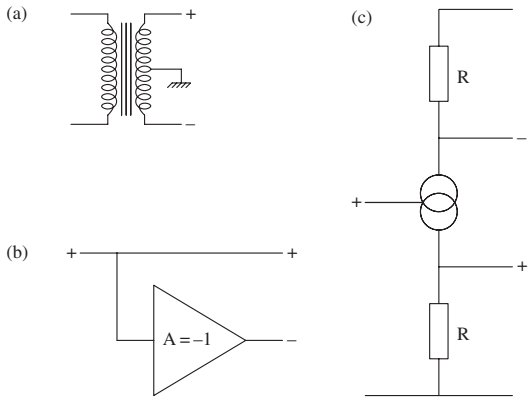


Fig. 6.15 *Fundamental basis of all phase splitters*

- We use a centre-tapped transformer in the same way that we use an output transformer to provide inverted and non-inverted signals. All of the previous considerations about transformers apply with a vengeance because of the comparatively high impedances involved, and the technique is not widely used, even though balance is near perfect under all conditions. See Fig. 6.15a.
- We have two outputs, one is the original signal, and the other is simply the input passed through an inverter. See Fig. 6.15b.
- A single device controls the flow of current in two resistors, one of which is connected to ground, and the other to HT. Increased current causes the instantaneous voltage dropped across each resistor to rise, so at any instant the voltage relative to ground is falling at one output, whilst the other is rising. See Fig. 6.15c.

All of the phase splitters employing the second method are based on the differential pair, whilst the third method is the basis of the concertina phase splitter. Triode phase splitters with low resistance outputs are very sensitive to their loading. They produce different output resistances when each output is loaded equally, compared to only one output being loaded. Phase splitters with a high r_a have an output resistance dominated by R_L, so phase splitters using pentodes and cascodes are immune to loading problems.

Loading sensitivity means that triode phase splitters should only ever be loaded by a stage that can be guaranteed to be Class A1 at all times if the balance of the phase splitter is not to be upset. (Any transition into Class B temporarily eliminates Miller capacitance at the input of the switched-off

valve, whilst Class A2 dramatically reduces input resistance of the switched-on valve once grid current flows.)

The differential, or long-tailed, pair and its derivatives

Most classic phase splitters were based on the differential pair, and much ingenuity was demonstrated in improving their performance.

A perfect differential pair comprises two devices (each with a load resistance) connected so as to allow a signal current to swing backwards and forwards between the load resistances without any loss whatsoever. Loss of signal current from the cathode to ground impairs performance, so the tail resistance is crucial, and should ideally be infinite.

The $R_k >> R_L$ solution

The differential pair can be optimized with a pentode or cascode constant current sink. An EF184 pentode can achieve a tail resistance >10 MΩ, and even the larger pentodes, such as the EL83, can manage 1 MΩ unaided. Performance at low frequencies can be improved by adding a transistor to make a hybrid cascode, but we will always be limited at high frequencies by C_{kh} from the differential pair's cathode, even if the sink is perfect. See Fig. 6.16.

The behaviour of the differential pair was discussed in Chapter 2. Provided that $R_k \approx \infty$, a balanced output must be achieved if $R_{L1} = R_{L2}$. Output resistance must also be identical from both outputs, and $r_{out} = r_a \| R_L$, as before.

However, if only one output is loaded, then:

$$r_{out} \approx \frac{R_L(R_L + 2\,r_a)}{2(R_L + r_a)} \approx \frac{R_L}{2}$$

So the stage should only be loaded by a stage that never strays out of Class A1.

The $R_k \approx R_L$ compensated solution

We accept that we cannot easily achieve a high tail resistance, and do not even try. We use a resistor, typically from 22 kΩ to 82 kΩ, as a tail, calculate what the errors will be, and try to correct them. This is known as the cathode coupled or Schmitt phase splitter.[4] See Fig. 6.17.

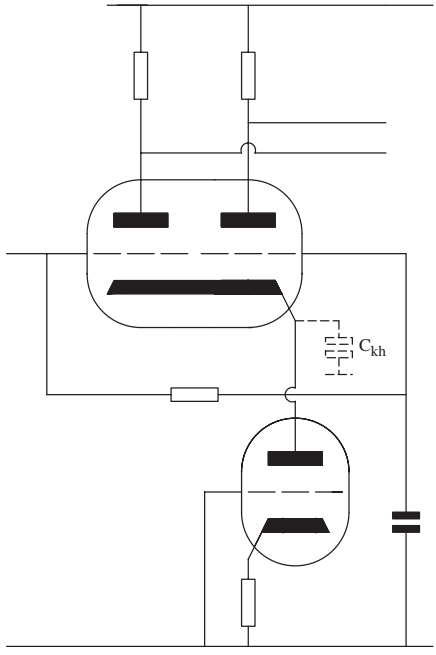


Fig. 6.16 *Differential pair with triode constant current sink as phase splitter*

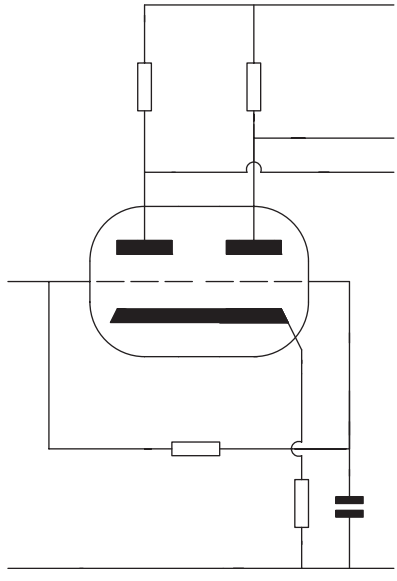


Fig. 6.17 *Cathode coupled phase splitter*

V_2 can be considered to be a grounded grid amplifier, fed from the cathode of V_1. It is this use of the first valve as a cathode follower to feed the second valve that results in the apparent loss of gain of the second valve, since for a cathode follower, $A_v < 1$. By inspection, we see that $2V_{gk}$ is required to drive the stage, so the gain of the compound stage to each output is half what we would expect from an individual valve.

If the outputs are in balance, then $V_1 = V_2$, so:

$$i_1 R_1 = i_2 R_2$$

The gain of V_2 is A_2, so:

$$v_{gk} = \frac{v_2}{A_2}$$

The signal current flowing in the cathode resistor is the out-of-balance output signal current:

$$v_{gk} = (i_1 - i_2) R_k$$

The signal at the output of V_2 must be:

$$(i_1 - i_2) R_k \cdot A_2 = i_2 R_2$$

Expanding and collecting I_2 terms:

$$i_1 \cdot R_k \cdot A_2 = i_2 (R_2 + R_k \cdot A_2)$$

Substituting $i_1 R_1 = i_2 R_2$, and simplifying:

$$\frac{R_2}{R_1} = \frac{R_2}{R_k \cdot A_2} + 1$$

This shows that unless the gain or the tail resistance of the stage is infinite, the ratio of the anode loads should be adjusted to maintain balance. Note that A_2 is the individual, unloaded, gain of V_2, and not the gain of the entire stage.

As an example, the Leak TL12+ phase splitter/driver was investigated. This uses an ECC81 and gives a gain for V_2 of 42 ($\mu = 53$, $r_a = 26.5\,k\Omega$), R_1 should therefore be 91 kΩ, and this is exactly what Leak used. See Fig. 6.18.

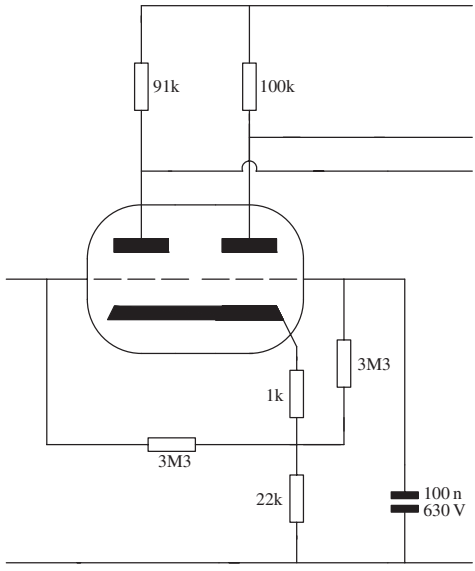


Fig. 6.18 *Cathode coupled phase splitter as used in Leak TL12+*

The output resistance of each half of the stage is slightly different because it is in parallel with a slightly different anode load, but curing this in order to preserve HF balance upsets the voltage balance at low frequencies. The nearest approximation that we can achieve is to include the grid leak resistances as part of the anode loads when calculating the necessary changes. The following grid leak resistors are 470 kΩ, so $R_{L2} = 100\,\text{k}\Omega \parallel 470\,\text{k}\Omega = 82.46\,\text{k}\Omega$, and the gain of V_2 falls to 40. The required total load for V_1 (including the 470 kΩ grid leak) is thus 75.7 kΩ, and $R_{L2} = 90.2\,\text{k}\Omega$.

LF balance is determined by the time constant of the grid decoupling capacitor and its series resistor, since it cannot hold the grid of V_2 to AC ground at very low frequencies.

Rather than tinker with resistor values whose calculation is critically dependent on valve parameters, the author would far rather add a constant current sink to the cathode to force the stage into balance.

The $R_k << R_L$ high feedback solution

We make no attempt at providing a large value of tail resistor, and rely on feedback to maintain balance; this circuit is known as the *floating paraphase* or *see-saw* phase splitter. Typically, this design uses a high μ valve, such as the ECC83, from whose data sheet this circuit was taken. See Fig. 6.19.

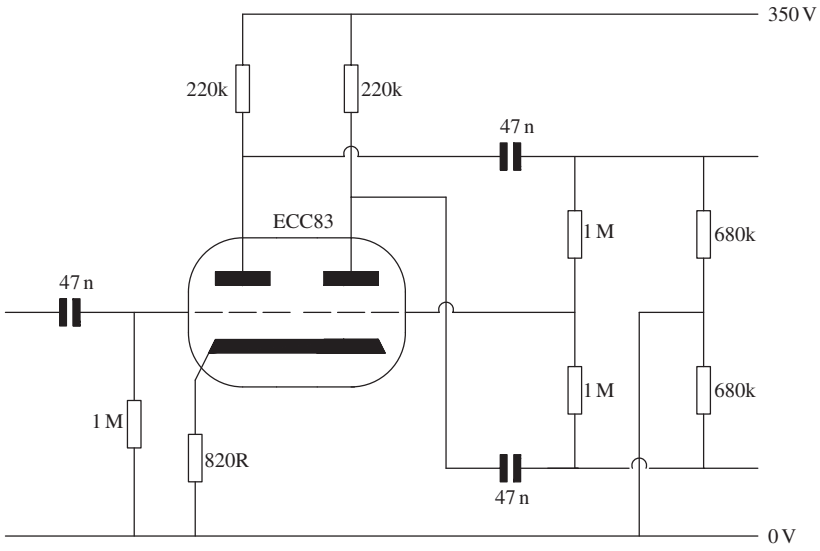


Fig. 6.19 *Floating paraphase or see-saw phase splitter (reproduced by courtesy of Philips Components Ltd)*

If we redraw this circuit, we see that V_2 is simply a unity gain invertor, whose gain is defined by resistors R_1 and R_2. See Fig. 6.20.

Since the open loop gain of V_2 is not infinite, these values must be adjusted to give a gain of −1. Unfortunately, the calculation is further complicated by

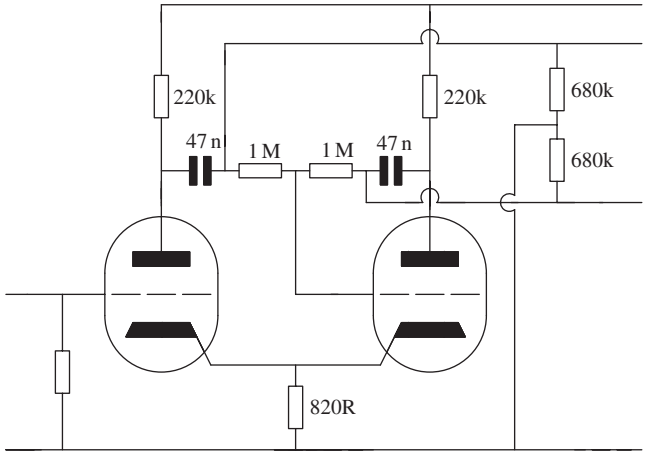


Fig. 6.20 *Floating paraphase phase splitter redrawn to reveal invertor*

the fact that R_2 affects the loading, and open loop gain, of the stage. V_2 also requires a build-out resistor to equalize its output resistance, which has been significantly reduced by negative feedback. Once these corrections have been made, the balance of this phase splitter is good, since the operation of V_2 is stabilized by negative feedback.

The circuit is analysed by first drawing a DC loadline to correspond to the 220 kΩ anode load. The Mullard operating point is at $V_a = 163$ V. The 1 MΩ feedback resistor is in parallel with this at AC, so we draw an AC loadline through the operating point corresponding to 180 kΩ. From this we find that the AC gain of the valve is 67.

We need to find the value of β that will give a final gain of 1, using:

$$\beta = \frac{1}{A} - \frac{1}{A_0}$$

We find that $\beta = 0.985$. The easiest way to achieve this is to increase the value of the feedback resistor:

$$R'_f = \frac{R_f}{\beta}$$

This gives a value of 1015 kΩ, so we would add 15 kΩ in series. So far, we have only discovered a 1.5% error, which is trivial, but if we consider the output resistances we find a much larger error. The output resistance of V_1 is r_a in parallel with R_L, and is ≈53 kΩ, but the output resistance of V_2 has been reduced by a factor of $(1 + \beta A_0)$, from 53 kΩ to ≈790 Ω. 52.2 kΩ of build-out resistance is therefore required, but the nearest standard value of 51 kΩ would be fine. These outputs are then each loaded by 680 kΩ, and if corrections are not made, the output from V_2 is ≈6% high.

In practice, these corrections were never made, which perhaps accounts for the poor reported performance of the stage. It might be thought that the common connection of the cathodes would improve balance, but V_2 has such heavy feedback that it can easily overcome any self-balancing action generated at the cathode.

The concertina phase splitter

The phase splitters based on the differential pair were all able to provide overall gain, but this was obtained at the expense of an output balance that was partially dependent on the matching of μ between the valves.

Although the concertina phase splitter does not provide gain, its output balance is almost totally determined by passive components, and valve characteristics hardly enter the picture. Conceptually, operation is very simple. Modulation of grid voltage causes a signal current to flow in the valve, the anode and cathode loads are equal and they have the same current flowing through them, so the signals generated across them are equal, implying perfect balance. See Fig. 6.21.

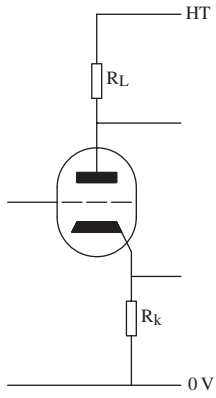


Fig. 6.21 *Concertina phase splitter*

Gain of the concertina

The gain of the concertina may be found using the standard triode gain equation, but noting that all undecoupled resistances down to ground via the anode resistance are multiplied by a factor of $(\mu + 1)$, so that:

$$r'_a = (\mu + 1)R_k + r_a$$

$$A = \frac{\mu R_L}{r_a + (\mu + 1)R_k + R_L}$$

But for the concertina, $R_k = R_L$, so:

$$A = \frac{\mu R_L}{R_L(\mu + 2) + r_a} \approx 1$$

Because of this low value of gain to the anode, Miller capacitance is very low, and the stage has wide bandwidth.

Output resistance with both terminals equally loaded (Class A1 loading)

The concertina is a special case ($R_k = R_a$) of an unbypassed common cathode amplifier with outputs taken from both cathode and anode. The general feedback equation is:

$$A = \frac{A_0}{(1 + \beta A_0)}$$

The denominator of the feedback equation is the factor by which resistances are changed. Since we know the gain of the concertina and the gain of a simple triode amplifier, we can substitute them into the feedback equation to solve for the feedback factor:

$$\frac{\mu R_L}{R_L(\mu + 2) + r_a} = \frac{\dfrac{\mu R_L}{R_L + r_a}}{\text{feedback factor}}$$

Cross-multiplying to find the feedback factor:

$$\text{feedback factor} = \frac{R_L(\mu + 2) + r_a}{R_L + r_a}$$

The anode output resistance of a common cathode triode amplifier with no feedback is:

$$r_{out} = \frac{R_L r_a}{R_L + r_a}$$

The feedback works to reduce anode output resistance, so this value must be divided by the feedback factor:

$$r'_{out} = \frac{R_L r_a}{R_L + r_a} \cdot \frac{R_L + r_a}{R_L(\mu + 2) + r_a}$$

The $(R_L + r_a)$ terms cancel, leaving:

$$r'_{out} = \frac{R_L r_a}{R_L(\mu + 2) + r_a}$$

Initially, it seems most surprising that series feedback ($R_k = R_a$, after all) should *reduce* output resistance from the anode so that $r_{out} \approx 1/gm$, but this can be understood by considering an external capacitive load on each output. In the same way that $R_k = R_a$ defines a gain of 1 at low frequencies, so $X_{C(k)} = X_{C(a)}$ defines a gain of 1 at high frequencies, and changing this ratio of capacitances certainly would change the gain or frequency response at high frequencies, since it would change the feedback ratio β.

Because $Z_k = Z_a$, the frequency response at each output is forced to be the same, so the output resistances must also be equal, and $r_{out(k)} = r_{out(a)}$.

Concertina output resistance, only one output loaded (Class B loading)

Looking into the cathode, we see R_k down to ground, in parallel with r_k the anode path to ground:

$$r_k = \frac{R_a + r_a}{\mu + 1}$$

Substituting:

$$r_{out(cathode)} = \frac{R_k \frac{R_a + r_a}{\mu + 1}}{R_k + \frac{R_a + r_a}{\mu + 1}}$$

Simplifying, and noting that $R_a = R_k = R_L$:

$$r_{out(cathode)} = \frac{R_L(R_L + r_a)}{R_L(\mu + 2) + r_a}$$

Alternatively,

$$r_{out(cathode)} = \frac{R_L + r_a}{(\mu + 2) + r_a/R_L}$$

In a practical application, the $(\mu + 2)$ term is usually significantly larger than the r_a/R_L term, so to a reasonable approximation:

$$r_{out(cathode)} \approx \frac{R_L + r_a}{\mu + 2}$$

Although the cathode output resistance can be calculated fully, the approximation is usually good enough, and generally gives a value of about 1 kΩ.

Looking into the anode, we see R_a to HT which is AC ground, in parallel with the cathode path to ground:

$$r'_a = R_L(\mu + 1) + r_a$$

Substituting:

$$r_{out(anode)} = \frac{R_L[R_L(\mu + 1) + r_a]}{R_L(\mu + 1) + r_a + R_L}$$

Tidying terms:

$$r_{out(anode)} = \frac{R_L^2(\mu + 1) + R_L \cdot r_a}{R_L(\mu + 2) + r_a}$$

If we inspect this equation closely, we see that the terms involving μ are the only significant terms, and that if μ is reasonably large, then $(\mu + 1) \approx (\mu + 2)$, so that $r_{out} \approx R_L$.

If there is a possibility of Class B loading, the concertina should use a low μ valve because this reduces $r_{out(anode)}$ and consequent imbalance, ECC88 and E182CC are good, but 6AH4 and 12B4-A are even better. It is usual to direct couple to the anode of the input stage, and let that determine the DC conditions of the concertina, resulting in the saving of a coupling capacitor and a low frequency time constant. Although the concertina is often criticized for its lack of gain, when directly coupled to the input stage, the combination generally gives almost double the gain of the same two valves used as a phase splitter based on the differential pair.

The input stage

The input stage is where global negative feedback is applied, so it must provide an inverting and a non-inverting input, both with low noise. The triode differential pair is an obvious candidate for this stage, but the common cathode triode or pentode can also be used, in which case, global feedback is applied at its cathode. See Fig. 6.22.

Design of the input stage is fairly trivial, but can be slightly complicated by the usual practice of direct coupling to the phase splitter, which restricts choice of anode operating conditions.

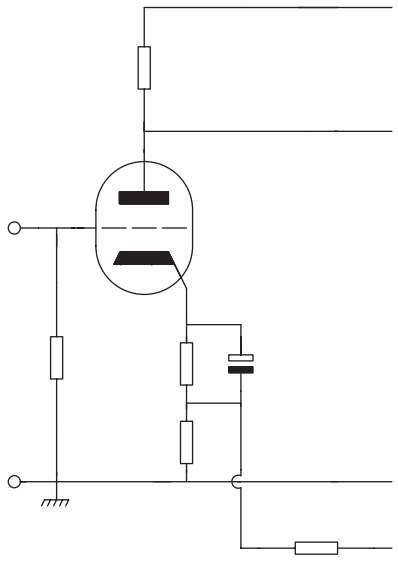


Fig. 6.22 *Application of global feedback at the input stage*

Stability

When we looked at *RC* networks in Chapter 1, we saw that a single *RC* network was asymptotic to 90° of phase shift. To make an oscillator, we need 180° of phase shift, so a single stage amplifier with one *RC* network causing an LF or HF cut-off *cannot* oscillate. If we cascade two such stages, we can approach 180° of phase shift, and if we feed this back into the input, it will ring, but not oscillate. If we have three such stages, it is a racing certainty that we can make the cascade oscillate when we apply feedback, and this is the basis of the phase shift oscillator.

To achieve oscillation, we need more than phase shift. Just because our feedback signal's phase has been shifted by 180°, it will not necessarily generate oscillation. We also need sufficient *loop gain*. The basis of oscillation is that it is self-sustaining; the gain of the amplifier must be sufficiently high to overcome the losses in the feedback loop before oscillation can occur. Loop gain is thus defined as the gain of the amplifier multiplied by the loss of the feedback loop.

If we have a phase shift of 180°, *and* loop gain ≥ 1, the circuit *will* oscillate.

Now that we have this definition, we can see how to avoid designing oscillators. We have two weapons at our disposal:

- We can reduce the number of stages, such that phase shift never reaches 180°. We rarely achieve this ideal, because the output transformer plus output stage plus driver stage harbours so many phase shifts, but the principle of minimizing the number of stages within a feedback loop is still valid.
- We attack the second condition of the oscillator statement, and reduce loop gain to ≤ 1 at the troublesome frequencies. This is the basis of all the methods that you will see for stabilizing amplifiers, and it is a powerful weapon capable of oppressing anything. Whether the resulting amplifier is of any use may be more debatable.

Slugging the dominant pole

This rather vibrant and mystifying description is actually very simple.

A pole in electronic jargon is simply another way of saying 'HF cut-off', just as a zero is an LF cut-off. What we aim to do is to make the amplifier *look* as if it only has one HF cut-off, which, with a maximum phase shift of 90°, is unconditionally stable. We look for the RC network with the lowest HF cut-off frequency, i.e. the dominant one, and we *slug* it with more capacitance to make it even lower.

Suppose that as a worst case, we cascaded four identical amplifiers, each with an HF cut-off frequency of 300 kHz, and a gain of 10. At 300 kHz, each amplifier contributes a phase shift of 45°, making a total shift of 180°. The gain of each amplifier is 3 dB down at 300 kHz, so the gain of each amplifier at that frequency is $10/\sqrt{2} = 7.071$, and the gain of the total amplifier must be:

$$A_{\text{total}} = \left(\frac{10}{\sqrt{2}}\right)^4 = 2500$$

In a typical amplifier we might want to reduce this gain from 2500 to 125, which would be 26 dB of negative feedback, and would reduce distortion to a twentieth of its original value. In order to do this, the feedback loop would have a loss of 0.0076. If we now check for stability, $0.0076 \times 2500 = 19$. The amplifier has a loop gain ≥ 1, *and* a phase shift of 180°, so it will oscillate.

We need to reduce the open loop gain at 300 kHz by a factor of 19, or 25.5 dB, to achieve stability (this has little effect on the frequency response of the final amplifier). Remembering that 6 dB/octave is equivalent to 20 dB/decade, reducing *one* cut-off from 300 kHz to 30 kHz will give us 20 dB reduction, and halving from 30 kHz to 15 kHz will give us another 6 dB, making 26 dB in total. This procedure may be formalized by the following statement:

The loop gain may be as large as the ratio of the two most dominant time constants.

To apply this rule, we simply choose how much feedback we want, calculate the loop gain, and adjust the dominant time constant until the ratio between it and the adjacent time constant is equal to the loop gain.

The amplifier is now stable, but only just, and it *will* ring. We should distance the dominant time constant still further to increase stability and remove ringing, certainly by a factor of 2, and preferably a little more. It should be realized that overzealous stability compensation reduces feedback, and compromises distortion reduction.

Many practical amplifiers, having exhausted the first two possible methods of achieving stability described, resort to manoeuvring the amplitude response independently of phase response using *step networks*. These must be adjusted on test, but if we are forced to use this method, computer-aided AC analysis is an excellent tool for eliminating poor possible solutions and narrowing in on the optimum solution. There are many analysis programs available, but don't be taken in by spurious precision – accuracy is only as good as the model that generated the calculations.

Realistically, we ought to consider that methods of achieving stability, in priority order, are:

- Reduce the number of time constants or stages within the loop.
- Slug the dominant pole (or zero).
- Fudge the phase/amplitude response using step networks.

There are some stability problems that are peculiar to valve amplifiers, and they have well-known symptoms and cures.

LF instability, or *motorboating*

This is an oscillation at about 1 or 2 Hz, and is invariably caused by unintentional feedback travelling through the power supply circuits, due to the rising impedance of filter capacitors at low frequencies. In effect, the entire amplifier becomes a relaxation oscillator.[5] The traditional cure was to insert

an LF step network, or to reduce the value of the coupling capacitors, *in the signal path*, so as to reduce the loop gain. This solution molests the second condition of the stability statement, but only treats the symptoms.

The *real* solution is to attack the first condition by *removing* the filter capacitors and their associated *RC* time constants by fitting HT regulators. This generally kills the problem stone dead. It is this improvement in stability that is the reason for the superior bass in designs using regulated supplies, since it removes previously unidentified LF ringing. It should be noted that this problem does not necessarily need a global feedback loop to make itself felt, and 'zero feedback' pre-amplifiers are not immune.

It is not unusual to discover that not only is the amplifier motorboating, but that it also has bursts of HF oscillation, known as *squegging*. If possible, it is best to cure the HF problem first, since it indicates marginal stability when the amplifier is under maximum stress and may be concealed once the LF instability has been cured.

Parasitic oscillation of the output stage and grid stoppers

The solution is almost encompassed by the description. Parasitics are the unwanted stray capacitances and inductances that result from the practical attempt to build an amplifier.

Miller capacitance in the valve combines with series inductance in the grid wiring to form a resonant circuit, so valves with a high *gm* (low r_k) are particularly prone to oscillation. (Inductance in the cathode part of the circuit is not a problem because it causes negative feedback which reduces loop gain.) The best cure is to damp the resonant circuit by fitting *grid stopper* resistors in series with the grid *as close as possible* to the grid pin of the valve base. The physical positioning of the resistor reduces the grid inductance, whilst a given value of resistance in the grid circuit is far more effective at increasing the loss of the resonant circuit without compromising frequency response than if it is placed in the cathode circuit. Remember that:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}}$$

Thus, Q is most dependent on series resistance, and adding 10 kΩ of series resistance to the grid circuit is a sure-fire way of oppressing parasitics.

For small-signal valves that are prone to this problem (E88CC, 5842, EC8010) a surface mount resistor actually touching the pin is ideal. Typical

values range from 100 Ω to 10 kΩ, and are usually found by experiment since individual layout is critical.

Parasitic oscillation of ultralinear output stages and g_2 stoppers

Ultralinear amplifiers with poor output transformers or paralleled pairs of output valves sometimes need a series RC network between anode and g_2. This is because this section of the winding has resistance in series with leakage inductance – the additional network attempts to return this impedance to a pure resistance. For 43% taps, the impedance between the anode and g_2 tap is ≈9% of the total anode to anode impedance, so this is a good starting point for the value of resistor, but both have to be determined empirically (adding them to both halves of the transformer simultaneously, which makes life difficult), and are often around 1 nF and 1 kΩ. Bear in mind that each capacitor has to withstand V_{HT} when the associated anode swings towards 0 V.

Stability of valve amplifiers is often described by the amount of additional feedback that would be required to cause oscillation. This is simply the ratio by which the dominant time constants were further distanced, over and above the necessary minimum required for stability. The Mullard 5-20 is proud to say that 10 dB more feedback would be required to cause instability, whereas the Williamson is questionable at LF even without additional feedback.

From this it should be realized that keeping a valve amplifier stable after applying global negative feedback is not a trivial task.

Classic power amplifiers

Now that we can recognize and analyse individual stages, we can investigate the design of some classic amplifiers such as the Williamson, the Mullard 5-20 and the Quad II.

The Williamson

The design of this amplifier was published in *Wireless World*[6] in 1947, and set a standard of performance that was years ahead of its time.

The input stage is the standard common cathode triode with 20 dB of global negative feedback applied from the loudspeaker output to the cathode. The phase splitter is a concertina circuit direct coupled from the input stage, and feeds a differential pair using both halves of a 6SN7. See Fig. 6.23.

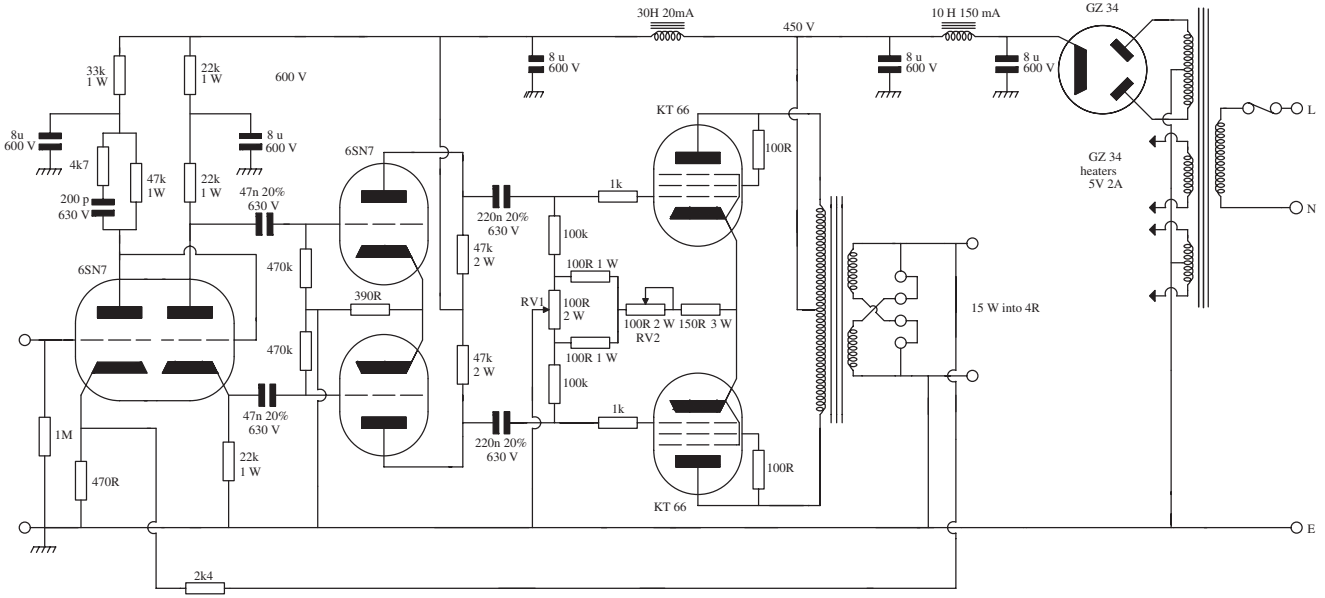


Fig. 6.23 *Williamson amplifier (by kind permission of Electronics World)*

The output stage is a push–pull pair of KT66 beam tetrodes operated as triodes that provide 15 W output in Class AB1, operating mostly in Class A. *RV*1 adjusts the DC balance of the output valves in order to minimize distortion due to the transformer core, whilst *RV*2 sets the quiescent current to 125 mA for the entire stage.

The linearity and headroom of each stage is excellent due to the careful positioning of operating points and choice of valves, but because this amplifier has four stages enclosed by the feedback loop, stability needs to be taken very seriously.

The input stage initially has an output resistance of $\approx 7.5\,\text{k}\Omega$, but this is raised by the feedback to $\approx 47\,\text{k}\Omega$. In combination with 12 pF of input capacitance from the concertina, this gives a high frequency cut-off of $\approx 280\,\text{kHz}$. However, this has been modified by adding the step compensation components *R*2, C1 to the anode circuit of *V*1. This circuit puts a step in the amplitude response which begins to fall at $\approx 130\,\text{kHz}$, but the phase response remains virtually unchanged until 280 kHz.

The concertina drives a driver stage with an input capacitance of 60 pF, and because the output transformer for the Williamson was very carefully specified, it seems unlikely that losses in the output transformer would cause the global feedback loop to force the driver stage out of Class A operation. The concertina thus faces a balanced load, and has an output resistance $\approx 350\ \Omega$, resulting in $f_{-3\,\text{dB}} = 7.5\,\text{MHz}$, which is sufficiently high to be insignificant.

The driver stage has an output resistance of $\approx 8.7\,\text{k}\Omega$, together with 55 pF of input capacitance from the output stage, the cut-off is $\approx 330\,\text{kHz}$, and the output transformer is specified to have a cut-off of 60 kHz.

The number of HF cut-offs within the feedback loop has not been minimized, and the dominant HF cut-off (the output transformer) is rather close to the pair which are next most dominant. Thus, the only remaining way to achieve stability at HF was to adjust the phase response independently of amplitude response by means of a step network.

At low frequencies it is more useful to consider time constants than –3 dB points. The input stage is direct coupled to the concertina, so we can ignore this. The concertina feeds the driver stage with a *CR* of $\approx 22\,\text{ms}$, as does the driver to output stage, and the output transformer is set to 48 ms. In view of this, it is not surprising that LF stability is questionable, as was conceded in the original *Wireless World* article. In 1952, Hafler and Keroes decided that their output stage would benefit from a Williamson driver,[7] they deliberately quintupled the concertina to driver stage coupling capacitors from 50 nF to 0.25 μF to separate the low frequency time constants. In the author's experience,

not operating the input stage and concertina from a shared HT can sometimes induce motorboating.

It should be remembered that in 1947, circuits were designed using long multiplication or tables of logarithms, and if speed was needed – slide rules. Computer-aided AC analysis was not an option! Most amplifiers were designed as carefully as possible, then adjusted on test for best response – and wide-bandwidth (>l MHz) oscilloscopes were recently developed luxuries.

The Mullard 5-20

This is a 20 W design[8] introduced by Mullard to sell the EL34 pentode. There is a great deal of similarity between this design, the Mullard 5-10 (10 W using EL84) and some Leak amplifiers. See Fig. 6.24.

The input stage is an EF86 pentode, which is responsible for the high sensitivity, but poor noise performance, of these amplifiers. Most of the cathode bias resistance is bypassed, since it would otherwise reduce the gain from around 120 to 33, which would be a waste of open loop gain that could be used to correct distortion produced by the output stage. Unadorned, the pentode has an output resistance of 100 kΩ, and drives ≈50 pF of input capacitance from the phase splitter, which would give a cut-off of 32 kHz, but this is modified by the usual compensation components across its anode load.

A slightly unusual feature is that the g_2 decoupling capacitor is connected between g_2 and cathode, rather than g_2 and ground. In most circuits, the cathode is at (AC) ground, and so there is no reason why the g_2 decoupling capacitor should not go to ground. In this circuit, there is appreciable negative feedback to the cathode, and so the g_2 capacitor must be connected to the cathode in order to hold g_2–k (AC) volts at zero, otherwise there would be *positive* feedback to g_2.

The cathode coupled phase splitter is combined with the driver circuit using an ECC83. When loaded by the output stage, for V_2, $A_v = 54$, but gain to one output is half this at 27.

The anode load resistors have not been modified to give perfect balance. With the 470 kΩ grid leak resistors of the output stage in parallel with the 180 kΩ anode loads, the *effective* anode load is 130 kΩ. Using the formula derived earlier, this means that V_{2b} should have an AC anode load 3% higher, and R_L for V_{2b} would then be 187 kΩ. Mullard did actually state this,[8] but probably assumed that most constructors would not have access to sufficiently high precision resistors to use the information.

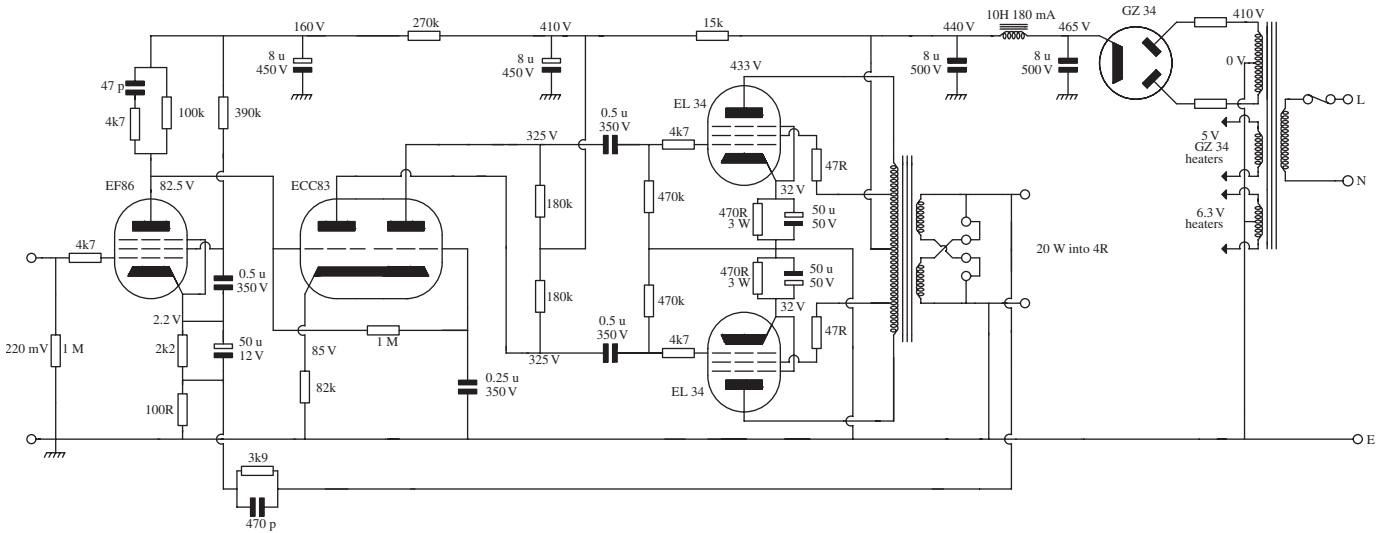


Fig. 6.24 *Mullard 5-20 (Mullard Ltd, originator of this design is now included in Philips Components Ltd)*

The output stage has an input capacitance of ≈30 pF, and the driver stage has an output resistance of 53 kΩ when loaded symmetrically, giving a cut-off at ≈100 kHz, which is quite poor. Loaded asymmetrically, the output resistance rises to ≈90 kΩ, which lowers the cut-off to ≈60 kHz.

Looking at the driver stage, we should investigate whether it is capable of driving the output stage. 85 V will be wasted across the 82 kΩ tail resistor, but with 410 V of HT, this still leaves us with 325 V. With the component values given, this puts the operating point at 240 V on the 180 kΩ DC loadline. Drawing the AC 130 kΩ loadline through this point shows that the stage would generate ≈4% 2nd harmonic distortion at full drive ($V_{out} = 18 V_{RMS}$), if it were not operated as a differential pair. Mullard claimed 0.4% distortion for the entire driver circuitry.

Although distortion appears satisfactory, the driver stage has only 10 dB of overload capability. When output stage gain begins to fall due to cathode feedback, or input capacitance loads the driver (particularly when the amplifier strays into Class B), or insufficient primary inductance in the output transformer, the global feedback loop will try to correct this by supplying greater drive to the output stage, and the 10 dB margin will be eroded, raising distortion.

The driver circuitry was designed to produce an amplifier of high sensitivity even after 30 dB of feedback had been applied, and this has forced other factors to be compromised. Whereas the Williamson sacrificed stability for linearity, the Mullard 5-20 achieves stability at the expense of linearity.

The output stage is a pair of EL34 in Blumlein configuration, with 43% taps for minimum distortion. Unlike the Williamson, there is no provision for adjusting or balancing bias, and this might seem to be a backward step.

Bias adjustment implies connecting the cathodes together and using a proportion of grid bias to provide the balance adjustment. Because the biasing is firmly set by the potentiometers, there is no self-regulation of bias current, and as the valves age, balance will need to be reset. In short, providing this adjustment ensures that it has to be used regularly.

By contrast, the Mullard 5-20 has separate cathode bias resistors and relies on automatic bias to hold the anode currents at their correct, and therefore equal, levels. In practice, this works well, although it does not quite achieve the low transformer core distortion of a freshly balanced adjustable system.

This system does have a disadvantage in that the individual cathode bias resistors apply series negative feedback to the output valves, raising their output resistance. The output transformer could be redesigned to maintain the match to the load, but this is undesirable as it would require a higher

primary to secondary turns ratio, which makes a high quality design more difficult to achieve. Because of this, the cathode bias resistors must be bypassed by capacitors, and this is where the problems really begin.

The capacitor is a short circuit to AC, and so prevents feedback, but its reactance rises at very low frequencies, so it is no longer a short circuit, and allows feedback. Because the output stage is matched to the load, feedback causes an immediate rise in distortion and reduction of output power due to the mismatch. The obvious solution is to fit a large enough capacitor to ensure that the LF cut-off for this combination is below all frequencies of interest, perhaps 1 Hz. Remembering that the resistance that the capacitor sees is R_k in parallel with r_k, we can easily calculate the value required.

For a pentode, $r_k = 1/gm$; a typical output pentode has $gm = 10\,mA/V$ at its working point, so $r_k \approx 100\ \Omega$, which is in parallel with a bias resistor of $\approx 300\,\Omega$, giving a total resistance of $75\,\Omega$. For 1 Hz, we therefore need $2000\,\mu F$ of capacitance.

$2000\,\mu F$ 50 V capacitors were simply not available at the time, and they weren't fitted. They are readily available now, but there are two reasons why you might wish to use a smaller value.

- A $2000\,\mu F$ capacitor has considerable inductance compared to its capacitance, and therefore allows feedback at HF. However, we can avoid this problem by using a low inductance electrolytic designed for use in switch-mode power supplies and bypassing it with a smaller capacitor.
- This reason is rather more subtle. If the output stage is driven into Class B by overload, each cathode tries to move more positively than negatively. It can't turn off any further, but it can certainly turn *on* harder. The cathode capacitor smooths these changes into a gently rising DC bias voltage, which biases the valve further into Class B, and the problem continues. The effect of this is that a momentary overload can cause distortion of following signals, even though they would normally have been within the capabilities of the amplifier. As the cathode bias capacitor becomes larger, the recovery time from overload lengthens. Theoretically, we never overload amplifiers, and this is not a problem, but occasional overload is inevitable, and the effects should be considered.

The ideal way to deal with all of these problems is to reduce the cathode bias resistor to $\leq 1\,\Omega$, so that it no longer causes noticeable feedback, and measure the current through it using an operational amplifier. This then feeds an asymmetric clipper so that when the valve strays into Class B and clips one half cycle, the clipper removes an equal amount from the other half cycle

before feeding the processed signal to an integrator. The integrator can have an *RC* time constant of almost any value we choose, and 10 s is not unusual. The output of the integrator is a smoothed DC voltage proportional to anode current, which can be compared to a fixed reference, and the difference between the two levels drives an amplifier whose output sets the negative *grid* bias for the output valve.

If the anode current of one valve is set as a reference, then the other valve, or valves, can share this reference, which then forces anode currents into balance. The increased complexity of this scheme is (partly) offset by its improved performance and reduction in HT voltage required, since the cathode bias scheme wastes HT. See Fig. 6.25.

This circuit was designed to sense a 40 mA anode current by developing 40 mV across the 1Ω resistor; the rest of the circuit is based on this 40 mV signal, so if a different current is to be sensed, the sense resistor should be changed to suit. The 5534 has a gain of 100, and amplifies the mean DC level to 4 V, with AC peaks rising to 8 V. Any peak above 8 V is clipped by the diode/transistor clamp, since the other half cycle will already have been clipped by the valve. The clipped signal is integrated by the 2.2 MΩ resistor in combination with the 470 nF capacitor, giving $\tau = 6.5$ s. The 071 compares this smoothed DC with a reference derived from the potential divider chain, and uses this to control the bias transistor. The clamp reference voltage set by the 2 kΩ variable resistor should be adjusted to achieve constant anode current under all conditions of overload. Although this circuit was designed to provide −11 V bias, this can easily be changed by returning the bias transistor's collector load to a more negative supply as necessary; no other changes are required.

The Quad II

The Quad II is an unusual design, which at first sight does not look too promising, but works because the design is synergetic.

In this design, not only has the phase splitter been combined with the driver stage, but it has also been combined with the input stage. In order to achieve the necessary gain, pentodes have been used. Output resistance is therefore high, as is input noise. To make matters worse, a variant of the see-saw phase splitter has been used. The output stage has local feedback, requiring increased drive voltage. See Fig. 6.26.

The output stage is a pair of KT66 beam tetrodes with anode and cathode loads split in the ratio 9.375:1. The cathode connection therefore provides little drive to the loudspeaker and may be considered to be series feedback

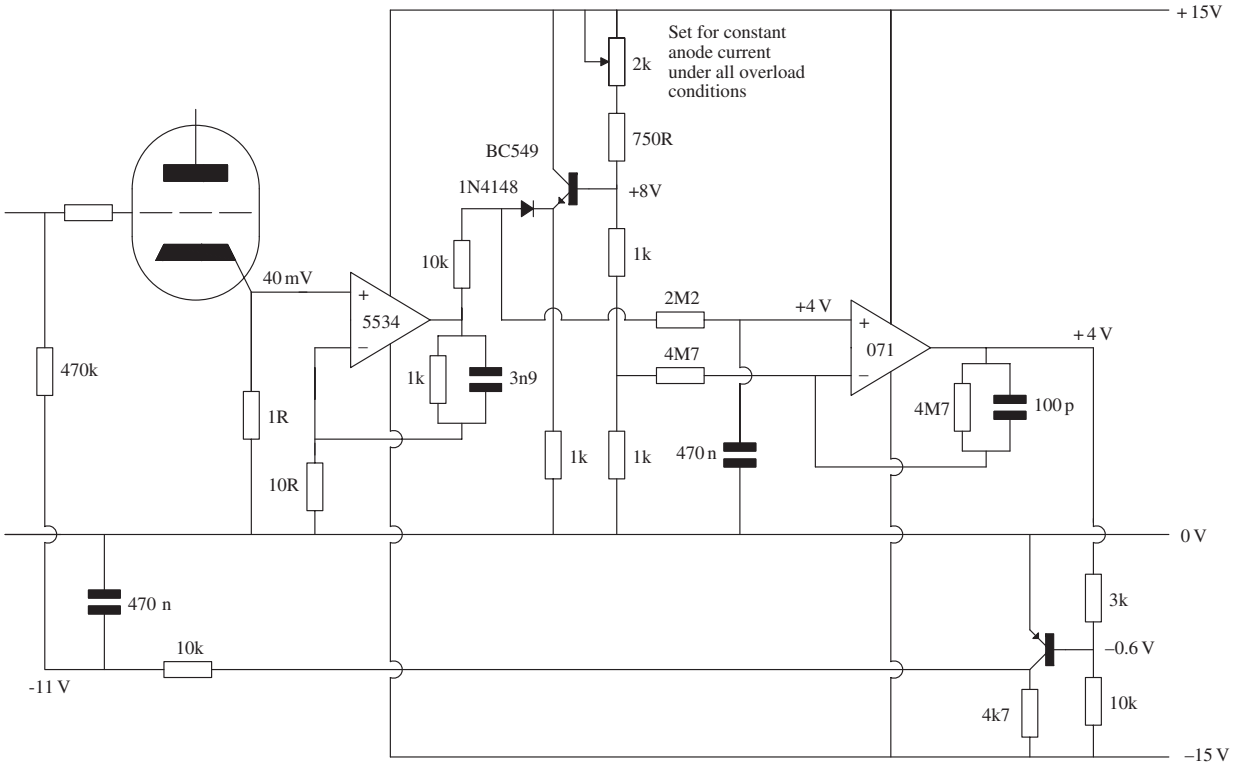


Fig. 6.25 *Principle of output bias servo*

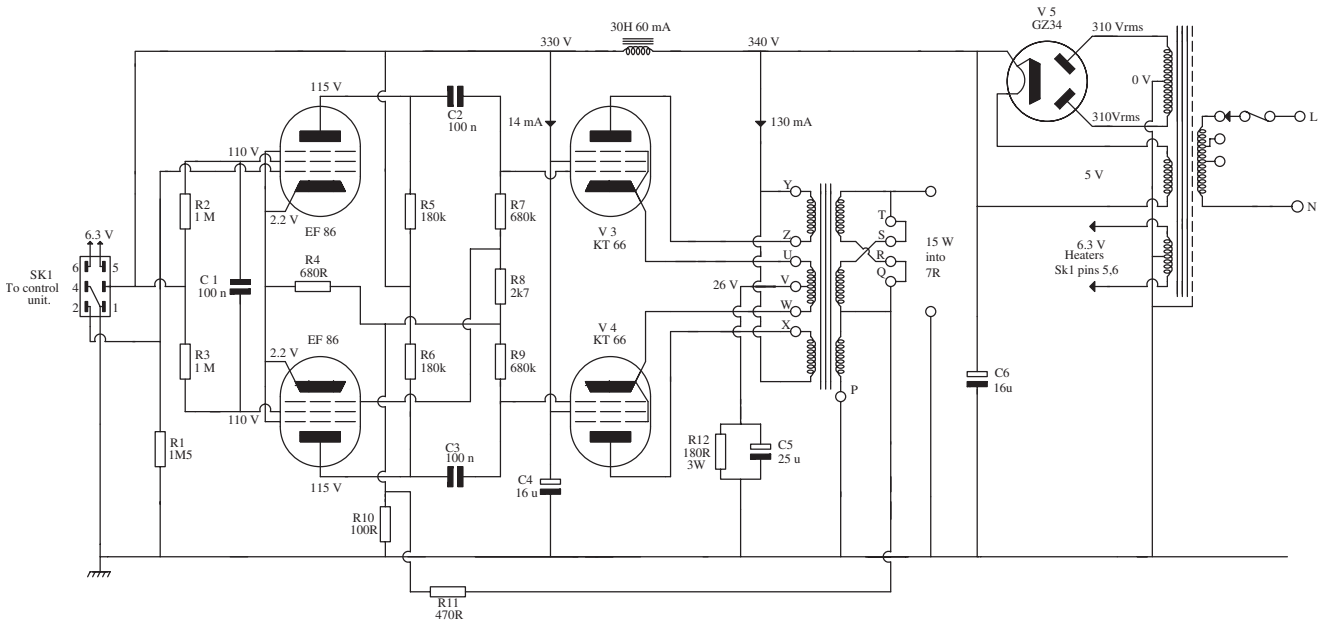


Fig. 6.26 *Quad II (by kind permission of Quad Electroacoustics Ltd)*

from the output transformer. However, the cathode current in the output transformer is the sum of the anode and g_2 currents, and it was found that this summation reduced 3rd harmonic distortion by a further 8 dB over that due to the negative feedback.[9]

The effect of this feedback on output resistance is the opposite to what might be intuitively expected.[10] If we simply leave a cathode resistor unbypassed, then this generates series feedback which increases r_a, whereas the transformer coupled feedback *reduces* r_a. This can easily be explained if we apply a short circuit as a load. Clearly, the output stage will be unable to drive any voltage into this load, but conversely, there will be no feedback signal applied to the cathodes. The grids will then be driven by the full input signal, rather than the input signal minus the feedback, so the output stage will be driven harder as it attempts to maintain its voltage into a short circuit. This action is directly equivalent to reducing output resistance, and the new value of output resistance can be found using the normal feedback equation.

The transformer primaries are equivalent to 3 kΩ anode to anode. With tetrodes, this low value of anode load results in a reduction of 3rd harmonic distortion, and an increase in 2nd harmonic, which is then cancelled by push–pull action in the output transformer (assuming that the output valves are perfectly matched).

The automatic bias is shared, so there is no provision for balancing anode current, and we can expect an increase in distortion at low frequencies due to saturation of the transformer core. Curiously, the cathode resistor was only rated at 3 W, yet it dissipates 3.8 W. If your Quad II distorts, a burnt-out cathode bias resistor may well be the cause.

Even with pentodes, there is not a great deal of gain from the driver circuitry, and input sensitivity is low; 1.4 V for full output. This is an excellent choice of input sensitivity for a power amplifier, as not only does it guarantee impeccable noise performance (even from a pentode), but it means that the input is far less susceptible to hum and noise from input cables or heater circuitry. The Quad II was only beaten in signal to noise performance by the Williamson, which was quieter because it had a triode input stage.

Despite being a variant of the see-saw phase splitter, the phase splitter/input stage does not rely on feedback for balance, and its operation is quite elegant. The output valves *must* each have a grid leak resistor, so instead of applying additional loading to the driver valves, a tapping is taken from one of these to provide the input for V_2. In theory, if this tapping has an attenuation equal to the gain of V_2, then the output of the phase splitter is balanced. Because of component variation, this will not always be true, so the cathodes of the two valves are tied together to improve balance.

Pentode stages have output resistance $\approx R_L$. Since R_L for the Quad input/phase splitter/driver is 180 kΩ, this would appear to be very poor at driving the ≈30 pF input capacitance of the output stage, resulting in a cut-off of ≈30 kHz. However, apart from the output transformer, this is the only HF cut-off in the circuit, and it is therefore not a problem. Each output valve requires a swing of $\approx 80\,V_{pk-pk}$, which is easily provided, because pentodes can approach 0 V more closely than triodes, and also because *LC* filtering was used on the HT line, rather than *RC* filtering, thus increasing the available HT. The *LC* filtered HT supply also feeds g_2 of the output valves, which has the valuable advantage of reducing hum, since the anode current of a tetrode or pentode is far more dependent on g_2 voltage than anode voltage.

Pentodes need to have g_2 decoupled to ground. Instead of each EF86 having a capacitor to ground, a single capacitor is connected between g_2 of the two valves. This has three advantages:

- If we had two individual capacitors, they would effectively be in series, with a centre tap to ground. Since each valve is connected to an equal but opposite signal, the centre tap would be at ground potential even if it were disconnected from ground. Therefore, we could cheerfully disconnect the centre tap from ground, leaving two capacitors in series that can be replaced by a single capacitor of half the value.
- Since this one capacitor is connected between two points of equal potential, it doesn't need the full voltage rating to ground. However, it is as well to consider the effect of fault conditions when determining the voltage rating, so this is not a great advantage.
- Connecting g_2 of each valve together at AC helps maintain balance in the same way as commoning the cathodes.

Although substituting one stage that combines the functions of input, phase splitter, and driver does not achieve the linearity of purpose designed stages, it achieves better linearity than the Mullard circuit because less gain is demanded from it.

With only a simple driver circuit and output stage within the feedback loop, the elegant Quad II has no stability problems.

New designs

We have investigated individual stages, we have looked at functional blocks, and we have seen how classic designs were configured. Rather than merely observing, it is now time to put that knowledge to use, and design an amplifier.

In previous editions of this book, it was suggested that an old amplifier could be cannibalized for its transformers and chassis. Sadly, this approach can no longer be justified because classic amplifiers are now likely to be 40 years old, and the amount of work expended in reusing components that might fail within 10 years is prohibitive. It is now cheaper and easier to make an amplifier with completely new components.

Single-ended madness

There are various reasons why a single-ended design has been included:

- The author feels that the single-ended genre should be given a fair trial and hanging. Or not, as the case may be.
- The author was given an NOS 6528, and realized that not only could it form the basis of a stereo amplifier, but that he already had a suitable mains transformer, chokes, rectifiers and HT capacitors. (The fact that a pair of custom-designed output transformers would have to be bought was not allowed to intrude upon this logic.)
- A single-ended amplifier is electrically simpler than push–pull, so it can make a good first project.

However, be warned. For a given output power, single-ended amplifiers are significantly bigger, heavier and more expensive than push–pull.

The 'Scrapbox Challenge' single-ended amplifier

Unfortunately for the transformer manufacturers, much of the ironmongery for this amplifier came from the author's salvage stock – hence the amplifier's name. Almost none of the components are critical, although alternatives will be offered as the design argument progresses.

Choice of output valve

Power amplifier design starts with required output power, which then leads to the choice of output valve(s). Thankfully, loudspeakers are gently becoming more sensitive as designers appreciate the advantages of a carefully chosen cone material, so even $\leq$10 W, suffices perfectly well without having to resort to expensive high efficiency designs such as horns. Having decided on

≤10 W, the next choice is *which* 300B to use. The world probably has enough 300B designs, so the 6528 made for an interesting alternative.

The Tung-Sol/Chatham 6528 (also made by Cetron and Raytheon) is a double triode intended for use as a series pass valve in regulated power supplies. The glass envelope resembles a GEC KT88, and it is internally akin to a 6080, but the detailed construction and consequent specifications are positively heroic.

	6080	**6528**
μ	2	9
gm	7 mA/V	37 mA/V
r_a	280 Ω	245 Ω
$P_{a(max.)}$	13 W	30 W
$I_{k(max.)}$	125 mA	300 mA
$V_{a(max.)}$	250 V	400 V
I_h	2.5 A	5 A
C_{ag}	8.6 pF	23.8 pF
C_{gk}	5.5 pF	17.8 pF

The primary attraction of the 6528 is its extremely low r_a, which suggests that we could use a low impedance output transformer – enabling better transformer design. We now need to decide what that primary impedance should be, so we will soon plot loadlines on the anode characteristics, but we must first clarify the class of output stage.

Choice of output class

Actually, there is very little room for manouevre. A single-ended amplifier can only be Class A. Class A2 implies grid current and requires power driving circuitry with extremely low output resistance, so this low power amplifier will be Class A1, where no grid current is permitted, allowing a simple voltage driver stage to suffice.

Choosing the DC operating point by considering output power and distortion

A power valve consumes expensive heater power, so it does not make sense to operate it at any point other than its maximum anode dissipation (30 W).

This means that our loadline must be a tangent to the $P_{a(max.)}$ curve. We also know that for maximum output, we should not clip one half of the waveform before the other, so our DC operating point must allow equal and opposite grid swings. Using the Tung-Sol/Chatham curves, sweeping a transparent ruler along the $P_{a(max.)}$ curve resulted in a 2 kΩ loadline with an operating point at $V_a = 255\,V$, which required $V_{gk} \approx -27\,V$, and $I_a = 120\,mA$. See Fig. 6.27.

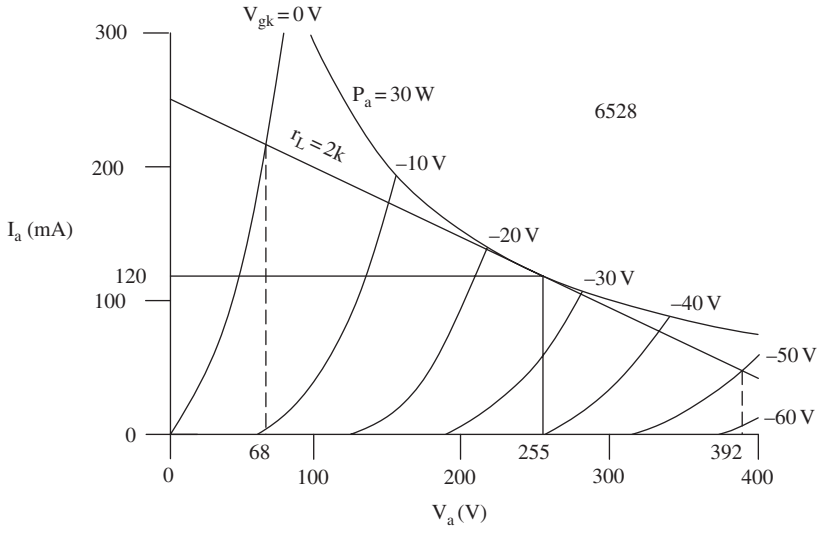


Fig. 6.27 *Operating conditions of 6528 output valve*

From the operating point, we can swing quite linearly to almost $V_{gk} = 0\,V$ before drawing grid current, and to a roughly equal and opposite swing of $V_{gk} = -50\,V$ before cut-off begins to cramp the grid curves. Once we know how many grid volts can be swung, we can check the corresponding anode swing. At $V_{gk} = -0\,V$, $V_a = 68\,V$, and at $V_{gk} = -50\,V$, $V_a = 392\,V$, so the peak-to-peak anode swing is 324 V. If we assume that we swing an undistorted sine wave, then $324\,V_{pk-pk} = 115\,V_{RMS}$.

The loadline passes from 500 V ($I_a = 0$) to 250 mA ($V_a = 0$), which corresponds to 2 kΩ, and this is much higher than the conventional choice of $2r_a$. If we know the anode load and the voltage swing across it, we can calculate the power dissipated in the load to see if the proposed loadline is acceptable:

$$P = \frac{V^2}{R} = \frac{115^2}{2000} = 6.6\ W$$

We can estimate the percentage of 2nd harmonic distortion using:

$$\%D_{\text{2nd harmonic}} \approx \frac{V_{\text{quiescent}} - \frac{V_{\text{max.}}+V_{\text{min.}}}{2}}{V_{\text{max.}} - V_{\text{min.}}} \times 100\%$$

At the chosen operating point, $V_{\text{max.}} = 392\,\text{V}$, $V_{\text{min.}} = 68\,\text{V}$, and $V_{\text{quiescent}} = 255\,V$, resulting in 7.7% 2nd harmonic distortion at full output.

This performance is typical for the genre, and other loadlines predicted significantly less audio power or worse distortion. (There is no point in being critical about the anode load, since real loudspeakers are nothing like pure, constant resistances anyway.)

Specifying the output transformer

We are now able to specify the output transformer:

Type : single ended
$I_{\text{DC}} = 120\,\text{mA}$
$P_{\text{max.}} \approx 6.6\,W$
Primary impedance $= 2\,\text{k}\Omega$

The transformer designer will immediately want to know the secondary load impedance, and because loudspeakers are *not* pure $8\,\Omega$ resistances, it is usually better to design for a $4\,\Omega$ load. Since a good output transformer has multiple secondary sections, transformer manufacturers commonly offer four sections that can be configured for $1\,\Omega$, $4\,\Omega$ (preferred setting for practical loudspeakers), $8\,\Omega$ and $16\,\Omega$ (ideal if you could find a genuine $16\,\Omega$ loudspeaker).

The next most important question is the lowest frequency for which maximum power is required. This is an expensive question. As an example of classic commercial practice, the Leak Stereo 20 and TL12+ amplifiers can only produce their specified power down to 50 Hz. The Sowter 9512 output transformer was specified for 8 W at 25 Hz in the unlikely event that practical power might exceed predicted power.

Biasing the valve

We could apply $-25\,\text{V}$ directly to the grid of the valve to set the required 120 mA of anode current, but a small drop in grid bias voltage would cause

$P_{a(max.)}$ to be exceeded instantly. This is the reason why valve manufacturers do not recommend grid bias with high mutual conductance valves – 35 mA/V is very high by valve standards.

We *must* use cathode bias. We need to drop 27 V across a resistor that passes 120 mA, so by Ohm's law:

$$R = \frac{V}{I} = \frac{27}{0.12} = 225\,\Omega$$

If the resistor has a voltage across it, and current passing through it, then it must dissipate power:

$$P = I \times V = 0.12 \times 27 = 3.24\,\text{W}$$

A 5 W power resistor is required as an absolute minimum. We could use an MPC-5 thick film resistor, which is non-inductive, but these get very hot when dissipating >2 W in still air, or we could screw a WH15 aluminium-clad wirewound resistor to the chassis, which would be cool, but inductance is a slight issue with low value wirewound resistors. After much havering, the prototype choice was a 200 Ω MPC-5 in series with a 100 Ω wirewound variable resistor, thus allowing a precise current to be set.

The cathode bypass capacitor

The cathode resistor must be bypassed with a capacitor to avoid unwanted feedback that would raise the valve's r_a, and require a new loadline. As in small-signal calculations, we look to see the AC resistances to ground from the cathode. Looking into the valve, the capacitor sees r_k:

$$r_k = \frac{r_a + R_L}{\mu + 1} = \frac{350 + 2000}{9 + 1} = 235\,\Omega$$

But this is in parallel with the 225 Ω cathode resistor R_k, so:

$$R_k // r_k = 115\,\Omega$$

In a small-signal stage, we would want the bypass to work down to 1 Hz, but this is not necessarily the case in a power stage. By definition, power stages swing large voltages and produce distortion. In a single-ended triode stage operated below clipping, the distortion is primarily 2nd harmonic, but this includes a DC component that the bypass capacitor integrates, shifting

the bias away from the intended DC operating point. Once the large signal has gone, the bypass capacitor gently recovers to the designed bias. Recovery time is determined by the time constant of the $f_{-3\ \mathrm{dB}}$ set by the capacitor in conjunction with $R_k//r_k$. If we chose 1 Hz as the $f_{-3\ \mathrm{dB}}$, this would imply a time constant, $\tau = 159$ ms. It takes 5τ for a CR combination to recover fully from a disturbance, and 0.8 s might be considered to be too long in musical terms, so we will set $f_{-3\ \mathrm{dB}} = 10$ Hz, which means that the output stage will recover from bias shift in only 80 ms:

$$C = \frac{1}{2 \times \pi \times 10 \times 115} \approx 1500\,\mu\mathrm{F}$$

The author didn't have any 1500 µF capacitors in stock, but he did have some low ESR 1000 µF 35 V, so this is what he used. Many engineering designs are done in this way – we calculate carefully (possibly on the back of an envelope), then choose the most suitable component from what is available at the time.

Finding the required HT voltage

The output transformer drops some HT across its primary winding resistance, so $R_{\mathrm{DC(primary)}}$ needs to be known. The transformer manufacturer can usually predict this value, but it is useful to have a measurement. For the transformer used, $R_{\mathrm{DC(primary)}} = 152\,\Omega$, so Ohm's law dictates that the voltage drop due to 120 mA of anode current is:

$$V = IR = 0.12 \times 152 = 18.24\,\mathrm{V}$$

Our design requires $V_a = 255$ V, with cathode bias of 27 V, and we drop 18 V in the output transformer, so we need an HT of 300 V at the top of the output transformer. Finding this HT voltage is significant because it determines the maximum HT available to the driver stage (unless we are prepared to add a subsidiary HT supply).

HT smoothing

Push–pull amplifiers cancel HT hum in their output transformer, but single-ended amplifiers are unable to cancel, so their HT supplies must be much quieter. Worse, single-ended amplifiers demand a changing current

(from $0 \leq I_{DC} \leq 2I_{DC}$) from their supply, so the output resistance of the supply is critical. Choke input supplies are therefore almost obligatory.

HT rectification

The heroic specifications of the 6528 were not achieved without an Achilles' heel. To avoid cathode stripping, the manufacturer specifically warns that the cathode requires 30 seconds to warm up before HT may be applied. This seems to be a perfect application for a valve rectifier. We only need 120 mA plus a little for the driver, perhaps 10 mA, and the HT voltage is only 300 V, so an EZ81 would be ideal.

In practice, typical valve rectifiers start conducting ≈10 s after power is applied, so a further delay is required, which can be provided by a thermal delay relay. Thermal delay relays look just like valves, and consist of a heater actuating a bimetallic strip in an evacuated glass envelope. The bimetallic strip is composed of two bonded metals having different thermal coefficients of expansion, so the strip bends when heated, and forms the moving half of a switch contact. Because the relay is almost evacuated, switch contact arcing is almost eliminated, and thermal losses are insignificant, so the heat required to make the strip actuate the switch contacts is determined purely by its specific heat capacity and mass. The delay can be increased by a factor of 3:1 over the relay's rated delay by reducing the voltage to its heater.

If the delay relay contacts are placed in the path of the rectifier's heater supply, then the delay of the relay is added to the delay of the rectifier, and the HT rises gently over a period of ≈5 s at the appointed time. Alternatively, many delay relays can safely switch AC mains or HT, but there must be negligible voltage between the heater and the moving contact. Unfortunately, the author didn't have any data for the Amperite 6N045T delay relay he found in his scrapbox, but from the part code he deduced that it required a 6.3 V supply, and that it would give a 45 s delay. The device was an all-glass envelope on a B9A button base, so it was easy to see which contact was which, and test the deduction. On test at 6.3 V, the heater drew 300 mA, and the switch contacts closed after 41 s.

When power is applied, the mains transformer delivers heater and HT voltage simultaneously to the valve rectifier, but its heater is cold, so the cathode suffers ion bombardment. Although placing the delay relay in the rectifier's heater circuit ensures that the HT to the audio circuitry rises gently from zero, it does mean that the rectifier suffers an extra 45 s of ion bombardment each time the amplifier is switched on. This is an engineering compromise – the EZ81 is a cheap sacrifice to appease the expensive audio parts.

The HT transformer

We need 300 V of HT at the top of the output transformer, and have elected to use valve rectification combined with a choke input supply. The choke drops voltage across its R_{DC}, so this must be determined. The author's scrapbox (more of a room, really) yielded a pair of 15 H 250 mA Parmeko chokes that looked hopeful and whose $R_{DC} = 136\,\Omega$, so 130 mA passing through one of these chokes would drop 17 V, and this would be added to the 300 V, to give 317 V.

Rather than calculate from first principles, we can use the rectifier manufacturer's choke regulation curves to determine required transformer voltage. Interpolation of Mullard EZ81 curves predicted $\approx 375\,V_{RMS}$ for our required $317\,V_{DC}$.

Further rummaging in the scrapbox uncovered a large C-core transformer with a pair of 375–0–375 V @ 250 mA windings, and numerous 6.3 V heater windings, so this seemed ideal, allowing dual mono construction.

HT choke suitability

We finally have sufficient information to confirm whether or not the posited 15 H 250 mA HT choke is satisfactory. Using equations from Chapter 5, and assuming 50 Hz mains:

$$i_{AC(\text{positive peak})} = \frac{v_{in(RMS)}}{1155L} = \frac{375}{1155 \times 15} = 22\,\text{mA}$$

$$i_{\text{total peak current}} = I_{DC} + i_{AC(\text{peak})}$$

$$= 130\,\text{mA} + 22\,\text{mA} = 152\,\text{mA}$$

Since the choke is rated at 250 mA, it should easily support this current. The minimum current required is:

$$I_{\text{min.(mA)}} = \frac{V_{in(RMS)}}{L_{(H)}} = \frac{375}{15} = 25\,\text{mA}$$

The output stage alone draws 120 mA, so we are safely above this lower limit.

We can estimate the hum due to HT once we know the proposed value of smoothing capacitor. The author had some 120 µF 400 V polypropylene capacitors in stock, so:

$$v_{\text{hum(RMS)}} = \frac{v_{\text{in(RMS)}}}{3.7 L_{(\text{H})} \text{C}_{(\mu\text{F})}} = \frac{375}{3.7 \times 15 \times 120} = 56\,\text{mV}$$

The anode load and r_a form a potential divider, so the ripple voltage seen at the anode is:

$$v_{\text{hum(anode)}} = \frac{r_a v_{\text{ripple}}}{r_a + R_L} = \frac{400 \times 56\,\text{mV}}{400 + 2000} = 9.3\,\text{mV}$$

An output transformer responds to the voltage *across* it, so it sees 56 mV − 9.3 mV = 47 mV of hum. At full output, the output stage swings 115 V_{RMS}, so 47 mV corresponds to a 68 dB signal/hum ratio, which is inadequate for high sensitivity loudspeakers, so a further stage of filtering is needed.

A second stage of *LC* filtering having a loss at 100 Hz of only 32 dB would improve the signal/hum ratio to 100 dB. 32 dB corresponds to a voltage ratio of 40, so the AC potential divider formed by the second *LC* filter would need $X_L/X_C \approx 40$. If another 120 µF capacitor were available, then even a 1 H 130 mA choke would be perfectly adequate.

However, the author *didn't* have another suitable pair of chokes, and had already realized that the amplifier was going to be large and heavy (even by valve standards). Adding yet more mass was not at all attractive.

The HT regulator option

LC filters might be good at reducing hum, but their output impedance is quite high (tens of ohms). This is particularly significant for a single-ended amplifier because the output valve is unable to distinguish between the reflected load of the loudspeaker through the output transformer and the output resistance of the supply. See Fig. 6.28.

The swing of the output valve is developed across both these components, yet we can only couple the swing developed across the output transformer. We lose power, and our output resistance rises. An HT regulator allows optimum bass performance from a single-ended amplifier.

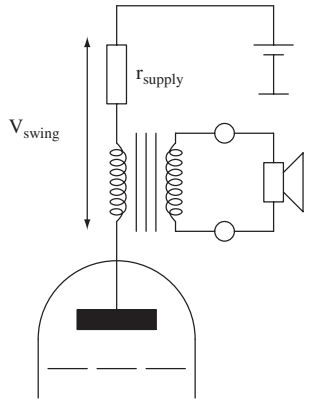


Fig. 6.28 *The effect of non-zero supply resistance on a power amplifier*

Each channel of the amplifier requires 300 V at 130 mA, and you could simply use the entire HT supply of Fig. 5.46 in Chapter 5 without any modifications. But we do not need a great deal of ripple rejection, so an adaptation of the simple two-transistor regulator is a possible alternative. See Fig. 6.29.

The two-transistor regulator has the advantage that it does not need to drop many volts, which reduces heat dissipation. We will assume that the regulator must drop $\geq$10 V, and that this will occur when the mains has dropped by 6% (as it is allowed to do). Thus, the nominal HT voltage required at the input to the regulator is:

$$V_{\mathrm{HT(nominal)}} = \frac{300\,\mathrm{V} + 10\,\mathrm{V}}{1 - 0.06} = 330\,\mathrm{V}$$

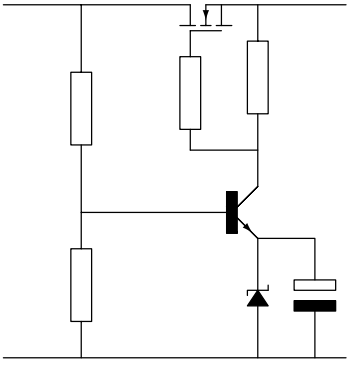


Fig. 6.29 *Two transistor HT regulator*

Checking the EZ81 data sheet, this would require a mains transformer with 412–0–412 V HT windings.

High voltage bipolar power transistors have rather low h_{FE}, they are slow, and they are expensive, so a high voltage MOSFET power transistor can often be a better choice for the series pass element.

When noise is not critical, it makes sense to make the Zener reference voltage as high as possible because this reduces dissipation in the error transistor, and also allows increased loop gain, which gives more feedback to reduce output resistance. 220 V is therefore a good Zener choice for a regulator that must provide an output of 285 V. Although 220 V Zeners are available (and the author had some in stock), three cascaded 72 V Zeners are better. The reason for this is that the very high voltage Zeners are rather noisy because they must be operated at a low current to reduce device dissipation ($P = IV$). Using three cascaded Zeners allowed a Zener current of 4 mA, which reduces noise. To reduce noise further, the Zeners are bypassed by a 22 μF 350 V capacitor.

The gate of the MOSFET will be at $V_{out} + V_{gs} = 300\,V + 4\,V = 304\,V$ (despite huge device variation, 4 V is a reasonably rough assumption for V_{gs} of a power MOSFET). Since the collector of the error transistor is connected to the gate of the MOSFET, and the emitter is tied to the 216 V (3×72 V) Zener reference, $V_{CE} = 304\,V - 216\,V = 88\,V$. We want the error transistor to pass 4 mA into the Zener reference, so $I_c = 4$ mA, and the power dissipated in the transistor is 352 mW. This is a significant calculation because it confirms that our choices of V_{CE} and I_c enable us to use a small-signal transistor.

When working, the error transistor only has $V_{CE} = 88\,V$, but at the instant of start-up, the 22 μF Zener bypass capacitor clamps the emitter of the error amplifier to 0 V, so it must be able to survive $V_{CE} = 330\,V$. The requirements for the error transistor are now clear, and the 400 V, 625 mW MPSA44 is ideal.

High voltage transistors have low h_{FE}, and the MPSA44 is no exception. When tested under the expected operating conditions, $h_{FE} \approx 100$. $I_c = 4$ mA, so $I_b = I_c/h_{FE} = 40\,\mu A$. Even if we pass 1 mA through the sampling potential divider chain, we cannot treat it as a pure potential divider because the 40 μA base current disturbs the result.

We set the sampling divider chain current by considering the power dissipation of the lower resistor first. If we use a 0.6 W component, and allow it to dissipate 0.2 W, it should remain sensibly cool. The resistor is connected to the base of the MPSA42, which is 0.7 V higher in voltage than the emitter, so the resistor has 217 V across it. Using $P = V^2/R$, its

resistance must be $217^2/0.2 = 235\,k\Omega$, so we will use the nearest preferred value of $240\,k\Omega$, which will dissipate 196 mW. The current through the resistor is $217\,V/240\,k\Omega = 904\,\mu A$.

Because the error transistor steals $40\,\mu A$ for its base, the upper potential divider resistor passes $904\,\mu A + 40\,\mu A = 944\,\mu A$ of current. The voltage across this resistor is $300\,V - 217\,V = 83\,V$, so its resistance must be $83\,V/944\,\mu A = 87.9\,k\Omega$, and the standard value of 91k will be fine.

There is no point in adding a capacitor across the upper resistor because the low attenuation of the divider chain (2.8 dB) means that it could only marginally improve ripple rejection, yet the required value would slow the response of the regulator to LF transient current demands (see Chapter 5).

The least critical circuit value to be calculated is the error transistor collector load resistance. We know that the input to the regulator is 330 V, and the collector voltage is 304 V, so this resistor has 26 V across it, so it passes $I_c = 4\,mA$, and its value must be $26\,V/4\,mA = 6.5\,k\Omega$, so the standard value of 6k2 will be just fine.

Back-of-an-envelope calculations suggested that the output resistance of this regulator would be $5\,m\Omega$, and that it would reject hum by >50 dB. In this instance, these figures are more than adequate, and offer better performance than another choke.

Determining amplifier output resistance

Checking AC conditions, r_a can be approximated by drawing a tangent to the −25 V grid line adjacent to the operating point, and this gives a value of $\approx 400\,\Omega$. This is significant because we can use it to calculate the output resistance of the amplifier. The output transformer matches the $4\,\Omega$ load of the loudspeaker to the $2\,k\Omega$ required by the valve, giving an impedance ratio of 500:1. Conversely, r_a of the valve is divided by this ratio, giving an output resistance of $0.8\,\Omega$. In order to allow loudspeakers to operate as their designers intended, the author would like to see a lower output resistance, but in comparison with many single-ended designs, this is quite low.

What are the driver stage requirements?

In the output stage, the maximum undistorted grid swing from the operating point is limited by the onset of grid current and a symmetrical swing in the opposite direction. Grid current occurs at 0 V, and by symmetry, the maximum opposite swing must be $2\,V_{gk}$, so the required peak-to-peak grid swing

for *any* Class A amplifier is *always* twice the grid to cathode bias. In our case, this means that we need $54\,V_{pk-pk}$ or $18\,V_{RMS}$.

We know that the 6528 must swing $\approx 115\,V_{RMS}$ at its anode and $\approx 19\,V_{RMS}$ at its grid, so it amplifies by a factor of $A_v \approx 6$, allowing us to find the Miller capacitance, which is $C_{ag} \cdot (A_v + 1) = 23.8\,pF \times (6 + 1) \approx 167\,pF$. This is in parallel with C_{gk} (17.8 pF), so the total input capacitance including strays is $\approx 200\,pF$.

We will investigate detailed arguments for the required high frequency response of an amplifier in Chapter 7, but if we make the sweeping assumption that $f_{-3\,dB} > 150\,kHz$, this requires a source resistance of:

$$r_s = \frac{1}{2 \times \pi \times 150\,000 \times 200 \times 10^{-12}} \approx 5300\,\Omega$$

The 6528 has a very high *gm*, and will oscillate locally at RF given half a chance. It therefore needs a grid stopper resistor to prevent oscillation, and the manufacturer's recommended value of 1 kΩ bites into our required source resistance, reducing it to 4.3 kΩ.

This is a low output resistance for a valve driver stage, and severely limits our design choice. In a practical common cathode driver stage, output resistance is roughly equal to the valve manufacturer's claimed value of r_a, so we are looking for a valve with a very low r_a. Frame-grid valves can achieve this value of r_a, but they tend to produce more 3rd harmonic distortion than valves with helical grids, so a driver stage using conventional valves with the output taken from a cathode would be preferable.

Driver stage topology

There are various options for the driver stage:

- A carefully designed common cathode stage could be DC coupled to a cathode follower. This could give low distortion and $r_s < 4.3\,k\Omega$.
- A μ-follower could give stunningly low distortion and $r_s < 4.3\,k\Omega$.
- An SRPP could give $r_s < 4.3\,k\Omega$, and a higher voltage swing than a μ-follower but with higher distortion.

The μ-follower is the obvious choice, but testing found that 290 V wasn't quite enough HT to enable the required $18\,V_{RMS}$ swing. Further, we know that the output stage is a capacitive load, for which the SRPP was *developed*,

and it can swing more signal than the μ-follower for a given HT voltage, so higher distortion is its only drawback.

Choice of valve for the driver stage

Ideally, we would like a driver valve that produces primarily 2nd harmonic distortion with insignificant higher harmonics, because it might conceivably allow some distortion cancellation with the output stage. Frame-grid valves are now almost eliminated (although the E88CC is easily one of the best), so the obvious choice is the *SN7/*N7 family. The author makes no apologies for arriving at the same choice as dozens of other single-ended designs. If sound engineering arguments dictate that round wheels are best, then that is what we will use.

However, an SRPP has its upper cathode at $^1/_2 V_{HT}$, so it requires an elevated heater supply if heater/cathode insulation is not to be strained. In a stereo amplifier, one *SN7/*N7 could be shared by the upper valves, and another for the lower. Alternatively, the 6J5 is half of a 6SN7, so a pair of these could be used in a monoblock amplifier. The author chose to use 6J5s because he had previously bought lots of them. Additionally, if individual valves are used, no unsightly metalwork is needed to modify the design later on, perhaps to use a high μ input valve with a medium μ second valve.

On test, the 6J5/6J5 SRPP stage easily swung $21\,V_{RMS}$ at 1 kHz, with 2nd harmonic at −40 dB, and 3rd at −54 dB. However, loading considerations meant that this measurement was taken purely by the oscilloscope/spectrum analyser, so the reliable measurement dynamic range was limited to only ≈55 dB, and higher harmonics could not be seen. Nevertheless, this distortion was felt to be tolerable compared to the predicted distortion of the output stage.

Determining the driver stage operating point

The title 'SRPP' (Shunt Regulated Push–Pull) implies that SRPP stages should use identical upper and lower valves. In practice, this seems not to be critical, and the author has not observed any measured advantage using different valves, but it is certainly easier to design with identical valves. The valves are in series (so they pass the same current), and identical, so the anode of the lower valve must be at half the HT voltage. Therefore, we design by considering the lower valve to be a common cathode stage with $V_a = {}^1/_2 V_{HT}$.

The 6J5 ideally likes $I_a \geq 8\,\text{mA}$ for constant r_a, or to consider it another way, $I_a \geq 8\,\text{mA}$ is likely to give lowest distortion. We will set $I_a = 8\,\text{mA}$, requiring $V_{gk} \approx 3.4\,\text{V}$.

As a corollary, driving 200 pF of shunt capacitance at 20 kHz with $19\,V_{RMS}$ requires $\approx 0.48\,\text{mA}_{RMS}$, or $\approx 1.3\,\text{mA}_{pk-pk}$ of signal current. An SRPP passing 8 mA should be comfortably able to provide this signal current without adding slewing distortion.

Setting driver stage bias

The upper valve of an SRPP *must* be resistor biased, *without* a bypass capacitor, otherwise there would be no signal to drive the valve, but the lower valve has a little more freedom.

The conventional cathode bias choice for the lower valve would be a 430 Ω resistor bypassed by an appropriately sized capacitor. However, when we designed the output stage, we considered the effects on bias after recovery from distortion. Since each half of the SRPP operates with only half of the available HT voltage (limiting signal swing), recovery after distortion or overload is important, so it would be better to use fixed bias in the lower valve. Fixed bias could be provided by grid bias or by semiconductor diode cathode bias. Grid bias is expensive, but semiconductor diode cathode bias can increase distortion. Fortunately, the author's measurements found that even in this stage, at these signal voltages, the additional distortion produced by LED bias was insignificant, and it allows instantaneous recovery from overload.

Is the output resistance and gain of the proposed driver stage adequate?

The output resistance of an SRPP driver stage would intuitively be expected to be sufficiently low, but this can be checked using:

$$r_{out} = \frac{r_{a_2}(R + r_{a_1})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

For this design, using measured values of $\mu = 21$ and $r_a = 7.1\,\text{k}\Omega$, the equation predicts $r_{out} \approx 2.3\,\text{k}\Omega$, which comfortably allows for a 1 kΩ cathode stopper resistor to reduce the likelihood of RF oscillation in the SRPP.

The gain of the SRPP stage is ≈14, and the output stage requires $\approx 18\,V_{RMS}$, so the input stage requires $\approx 1.3\,V_{RMS}$ to drive the amplifier to

full output – which is fortunate, because it allows ≈3 dB of gain from a standard 2 V_{RMS} CD player to allow for poorly mastered recordings.

But what about feedback?

It is de rigeur for single-ended valve amplifiers not to use global negative feedback. The argument is that their distortion is primarily 2nd harmonic, innocuous, and proportional to level, so feedback would translate distortion harmonics up in frequency to where they are more noticeable. This argument is plausible provided that the distortion is <5% at full output power, because historical tests of distortion could not detect 2nd harmonic distortion below 5%. Sadly, these tests were made even before early low distortion loudspeakers such as the Quad ESL57 full-range electrostatic became available, so their current validity is distinctly questionable.

However, a small amount of local feedback can be applied at the output stage by including the secondary of the output transformer in the cathode circuit.

Summing up

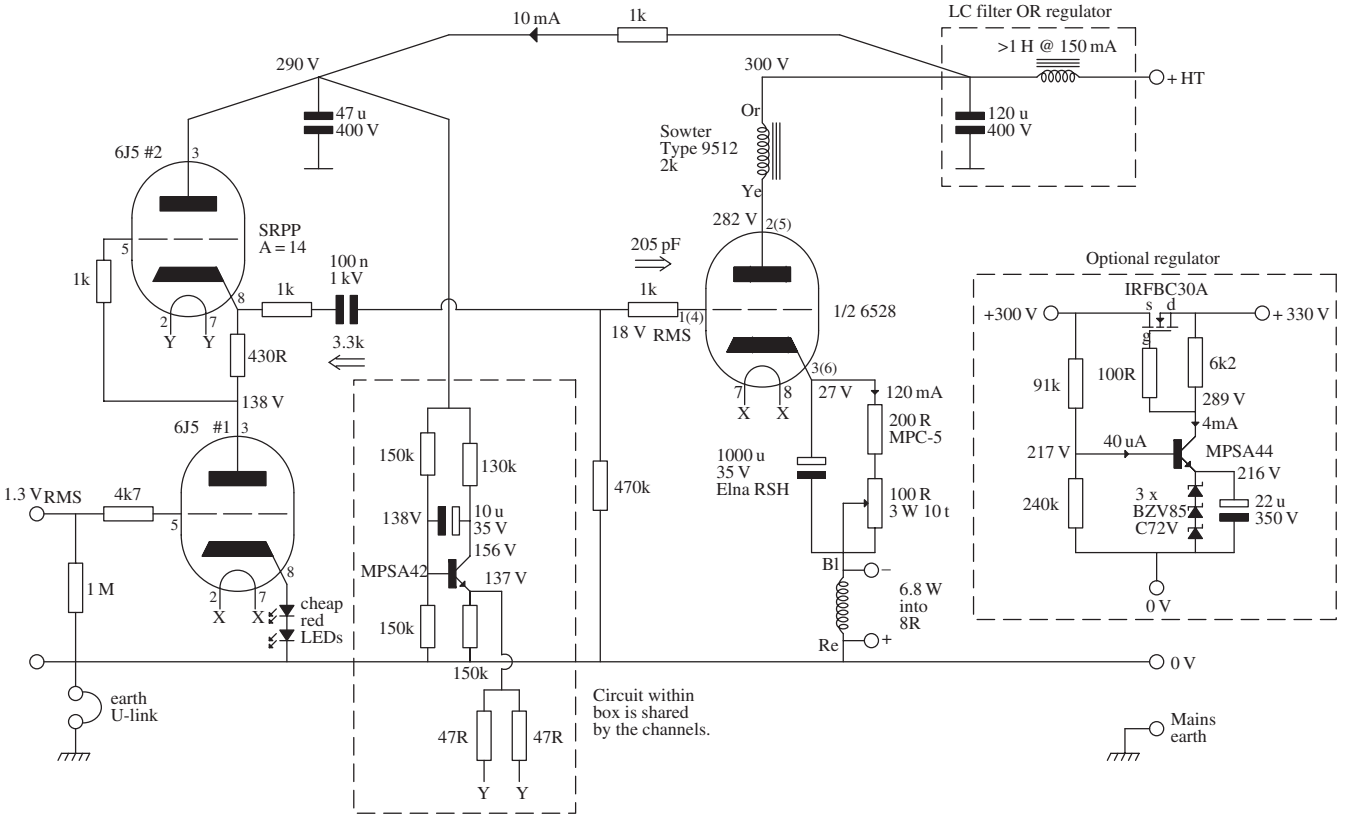
Now that detailed design is complete, it is worth reviewing the consequences of the design choices to see whether the whole design looks worthwhile. See Fig. 6.30.

Predicted output power ≈6 W with ≈8% distortion
Input sensitivity for full power ≈1.3 V_{RMS}

The amplifier was built to test the predictions and weighs 64 lb. Put another way, it weighs ≥10 lb per stereo watt. Compared to a push–pull design, it is heavy and expensive to achieve a limited objective, but precisely the same allegation would be levelled by a semiconductor designer at *any* valve amplifier. Valve amplifiers are the steam engines of the electronic world – and they arouse similar passion.

Teething problems

Racing car engines achieve their outstanding performance by operating every part *just* under its limit – so small errors are catastrophic. The 6528 can be considered to be a racing car engine in that it is rated at 30 W per anode, and


10 mA
1k
LC filter OR regulator
>1 H @ 150 mA
+ HT
290 V
300 V
47 u
400 V
120 u
400 V
Or
Sowter
Type 9512
2k
Ye
6J5 #2
SRPP
A = 14
1k
100 n
1 kV
205 pF
282 V
2(5)
1/2 6528
Optional regulator
IRFBC30A
+300 V
+ 330 V
s
d
g
91k
100R
6k2
289 V
4mA
217 V
40 uA
MPSA44
216 V
240k
3 x
BZV85
C72V
22 u
350 V
0 V
1k
3.3k
430R
138 V
18 V RMS
1k
1(4)
3(6)
7
8
X
X
27 V
120 mA
200 R
MPC-5
100 R
3 W 10 t
1000 u
35 V
Elna RSH
Bl
–
6.8 W
into
8R
Re
+
6J5 #1
1.3 V RMS
4k7
1 M
cheap
red
LEDs
150k
130k
138V
10 u
35 V
156 V
MPSA42
137 V
150k
150k
470k
47R
47R
Y
Y
0 V
earth
U-link
Circuit within
box is shared
by the channels.
Mains
earth

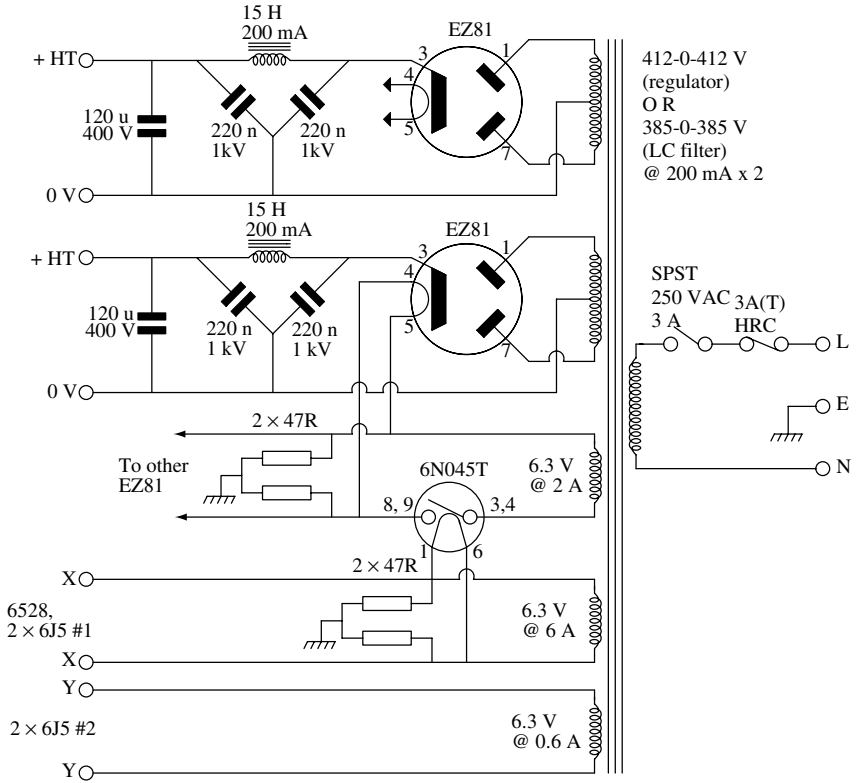


Fig. 6.30 *Practical 'Scrapbox Challenge' power amplifier*

consumes 31.5 W of heater power, so it has to lose 91.5 W of heat from an envelope the size of a KT88 ($P_{total} = 52$ W). Initial testing therefore concentrated on ensuring that the 6528 would not expire before the chequered flag.

The 6528 valve base was mounted on a wire finger-guard intended for an 80 mm fan, and an 80 mm super low noise fan blows gently from below the valve base. As a consequence, even though the valve is operating at maximum P_a, its measured envelope temperature is just within limits and the chassis is stone cold.

At the first test, the AC heater voltage at the valve base of the 6528 measured by a reliable true RMS meter was 6.5 V instead of 6.3 V, so the mains transformer primary tapping was changed from 240 V to 250 V, which reduced the heater voltage to 6.296 V – which *is* close enough. (Because the AC mains waveform contains considerable distortion, AC heater voltage measurements should always be made by a true RMS meter having good accuracy up to 1 kHz.)

Additionally, one of the 6J5 had to be rejected because of poor heater/cathode insulation (despite measuring $>25\,M\Omega_{(hot)}$ on the Avo VCM163 valve tester). Even though the heaters are decoupled to ground, the fault caused 5 mV of hum and rectifier switching spikes to be fed to the output stage.

Listening tests

Because of the importance of the trial, the 'Scrapbox Challenge' amplifier was carefully auditioned over a considerable time through the author's pair of LS3/5a. Although the LS3/5a is a very nice little loudspeaker, it is not ideally suited to weedy amplifiers. Unfortunately, they are the only passive crossover loudspeakers with any pretensions to quality that the author owns.

The amplifier started barely tolerably, but improved greatly over the first four hours of listening, and the fact that a bottle of Veuve Clicquot was symbolically opened seconds after the music began is quite irrelevant.

Designer's observations

The amplifier's wiring was completed over a weekend followed by two national holidays. Murphy's Law thus dictated that missing parts would only be discovered late on the Friday evening. The author intended to use his variant of choke snubbers for the HT supply, but the cupboard was

inexplicably bare of 220 nF polypropylene film/foil, so he was initially forced to use a traditional 10 nF film/foil + 10 kΩ snubber (although this defect was corrected a week later). With hindsight, this was fortuitous, because the poorer snubbing revealed that the cheap EZ81 rectifier switches on and off surprisingly cleanly and provoked very little ringing. See Fig. 6.31.

Building the 'Scrapbox Challenge' amplifier has reinforced the author's deepening suspicion that the magnetic cores of chokes and transformers can seriously deteriorate with age. The predicted ripple after the first stage of HT filtering was 56 mV_{RMS}, but the measured value was 7 dB higher at 124 mV_{RMS}. The 120 µF capacitors were measured and cleared, so the finger of suspicion points at the chokes (one of which had previous form, since it was guilty of buzzing on an early EL84 amplifier), and it seems likely that their inductance may have fallen due to deterioration of the magnetic core. Additionally, the oversize C-core mains transformer throbbed even before HT current was drawn, and required a 5 A fuse in the mains plug just to survive the inrush current as the amplifier was switched on. Moral: 40-year-old

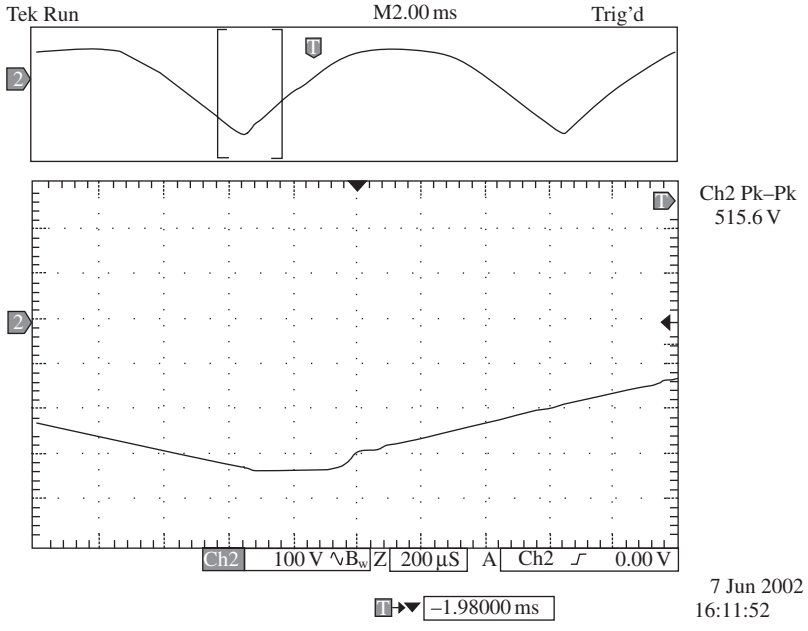


Fig. 6.31 *Voltage waveform at the output of the EZ81 rectifier with traditional snubber. (The lower, expanded trace shows rectifier behaviour as it switches on and off)*

electromagnetic components could be the weakest link. A measured failing *might* not be due to your design.

Listening closely at switch-on (before the delay relay activated the HT rectifiers) revealed that some hum was being induced directly into the output transformers from the mains transformer. If you decide to build this amplifier, it would be best to build it on two chassis, one for the audio circuitry and one for the power supply – allowing the noisy power supply to be distanced from the sensitive amplifier.

Conclusions

With the right programme, this amplifier is extremely easy on the ear. It *isn't* accurate – the inevitable high output resistance (2.1Ω on 8Ω setting) causes underdamped bass, and it falls apart on choral music. It is very heavy (11 lb/W) compared to (3.5 lb/W) for the Bevois Valley amplifier, and is almost twice the size. Nevertheless, it is not an amplifier that the author would like to lose.

On test, the amplifier delivered 6.8 W at 1 kHz with 3.2% THD + N. The amplifier was tested with and without cathode feedback at 1 dB below full power.

	2nd	3rd	4th	5th	6th
Without cathode feedback	−25.2 dB	−57.8 dB	−55.6 dB	−60.6 dB	−58 dB
With cathode feedback	−31.7 dB	−64.3 dB	−57.1 dB	−74.3 dB	–
Improvement	6.5 dB	6.5 dB	1.5 dB	13.7 dB	–

The feedback reduced amplifier gain by 3 dB, yet the table shows a significantly greater improvement on all harmonics except the 4th – making it a worthwhile trade.

On the available evidence, the case has not quite been proven, so the hanging has been postponed, even though the amplifier's distortion figures are risible compared to a competently designed transistor amplifier.

Single-ended amplifiers are very definitely a matter of personal taste.

Push–pull sanity

Our next design will be a 10 W Class AB1 push–pull amplifier using EL84 output valves in Blumlein or 'ultralinear' configuration.

There are two reasons for this choice:

- It is cheap. If we need ≈330 V HT supply, this can be smoothed by 385 V capacitors intended for switch-mode power supplies, and the HT could be provided by an isolating transformer with a silicon bridge rectifier and reservoir capacitor. If any mistakes are going to be made, then it is best to make them with reasonably inexpensive components rather than expensive ones.
- Powerful amplifiers require considerably more skill in layout and construction, and generate bigger bangs – they are not really recommended for the novice constructor.

The 'Bevois Valley' amplifier

This design acquired its name because the prototype was built from a pair of mono amplifiers bought for £15 (including pre-amplifier) in Bevois Valley. Sadly, the junk shop that sold these amplifiers closed (permanently) shortly after their sale in 1982.

As was noted earlier, once the output valves have been chosen, the output transformer configuration is limited, and therefore the entire output stage is fixed. We will use a transformer with a primary impedance of ≈8 kΩ anode to anode, having 43% taps for minimum distortion. It will need an HT voltage of 320 V, and reference to the Mullard EL84 data sheet shows that each valve will require $8\,V_{RMS}$ for full output.

Our task is to design superior driving circuitry. To do this, we will need to define what we need, so we will write a specification. This will enable clear thinking, and the electronic design will follow on naturally:

- *Low noise*. With the low noise obtainable from CD or a good LP, we really can't afford to have *any* noticeable noise in the power amplifier. 100 dB S/N, relative to full output power, is not an unreasonable figure to aim for, so this rules out pentodes and high sensitivity.
- *No hum*. This is the new millennium, not the 1920s! Hum is not acceptable under any conditions. This implies superb standards of construction, and/or DC heaters for the input stage.
- *Stability*. This might seem obvious, but some commercial designs that the author has endured were a little relaxed about this. Good stability requires an absolute minimum of stages.
- *Distortion*. This is a tricky topic. If you want distortion measured in ppm (parts per million), then you had better buy a decent transistor amplifier. If

you think that hearing is everything and measurement is nothing, then sell the house and build a huge single-ended triode amplifier. We have to be honest about this. Valve amplifiers do not always measure well, but they *do* sound good. Presumably, we listen to music to enjoy it, so this quality is important.

- *Simplicity*. Valve designs should be simple. Simple systems tend to have simple shortcomings. Additionally, they are repairable. Complex systems are built on silicon, have lots of legs, are repeatable and disposable.

Put together, these criteria demand that we use a concertina phase splitter direct coupled from the input stage without a driver stage, and we can instantly draw a circuit diagram. That this circuit is quite similar to the GEC912+[11] demonstrates that there is little new under the sun. See Fig. 6.32.

Since we are driving the output valves directly from the phase splitter, linearity of the phase splitter is paramount. Our chosen phase splitter only has a gain of ≈ 1, so the input stage will also need excellent linearity. Only three valves are really suitable for a concertina stage with limited HT current: *SN7/*N7, ECC82, E88CC. The *SN7/*N7 family needs more HT than we have available, and the ECC82 distorts. We will use the E88CC.

Optimization of DC conditions

Because the two stages are DC coupled, the design of the two stages is interactive. As before, the way to deal with an awkward problem like this is to garner as many facts as possible, label the drawing, and see if anything useful appears. Having chosen a concertina stage, we can start by labelling the anode and cathode loads as $22\,\text{k}\Omega$. (This traditional value is used because $R_{\text{out}} \approx R_{\text{L}}$ if loaded by a stage entering Class B, and whilst we need to minimize output resistance, a significantly lower value would result in excessive power dissipation.) See Fig. 6.33.

Previously, when we used the E88CC, we observed that linearity was best when $V_{\text{gk}} = -2.5\,\text{V}$, this tends to happen at $V_{\text{a}} = 80$ to $90\,\text{V}$. Although the concertina operates under heavy feedback, it would be preferable if it were linear *before* feedback. So we want to juggle conditions such that *both* valves are biased with $V_{\text{gk}} = -2.5\,\text{V}$.

The only way of doing this is by an iterative process. We know that both stages will have an HT of $\leq 300\,\text{V}$ since we have to drop volts from the output stage. We also know that each stage will have $V_{\text{a}} = 80$ to $90\,\text{V}$, at

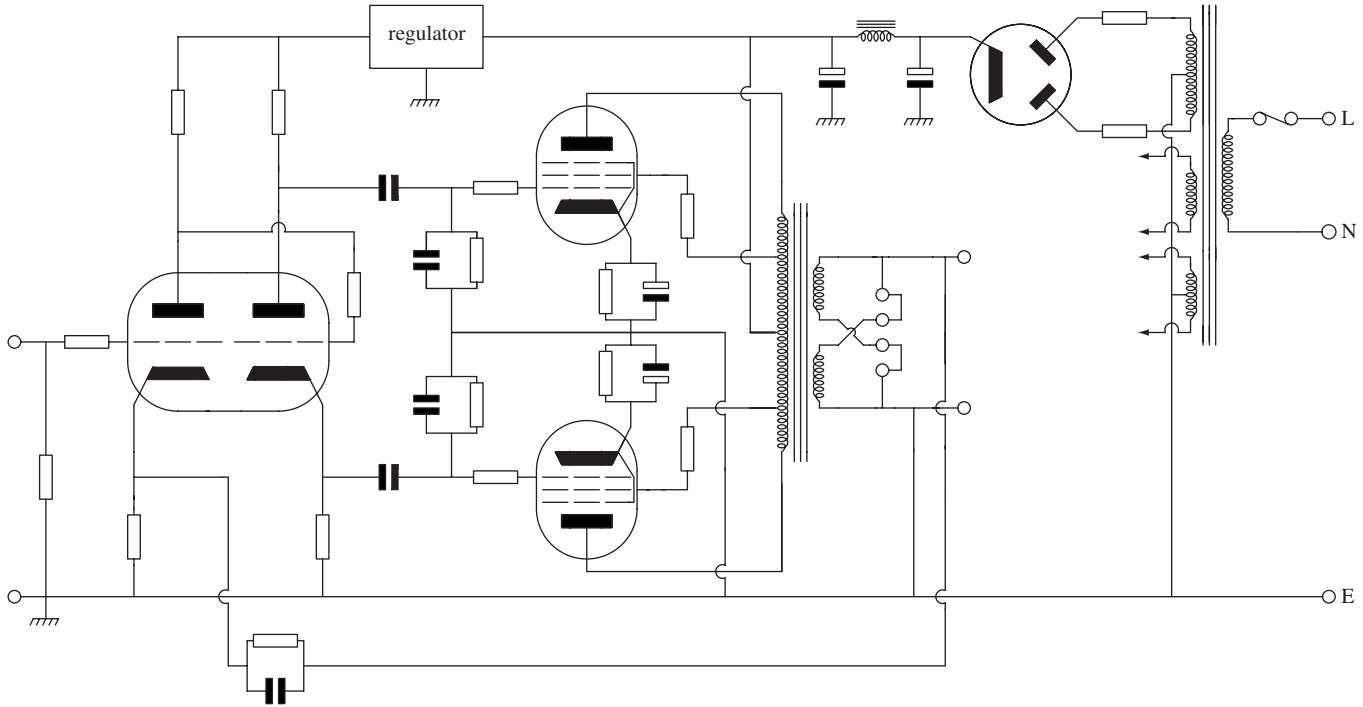


Fig. 6.32 *Power amplifier using concertina phase splitter*

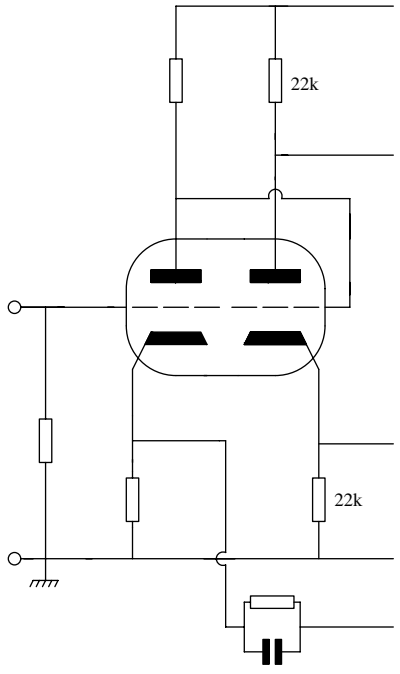


Fig. 6.33 *Determining the operating conditions of the driver/input stage*

$V_{gk} = -2.5\,V$. We draw the loadline for the concertina first and find V_a for $V_{gk} = -2.5\,V$. This value is then subtracted from the HT voltage to give the voltage across R_k and R_a, and divided by 2 to give the voltage across R_k. The voltage on the grid will be 2.5 V lower than this, and this will be equal to the anode voltage of V_1. We now draw a loadline for V_1 and see if the optimum anode voltage corresponds with the voltage we have just derived. If it doesn't (and it almost certainly won't) our only variable is HT voltage. Fortunately, a few iterations found that an HT voltage of 285 V met the requirements, and this will be provided by a regulator.

Ideally, we would also juggle the signal current in the input stage to be equal and opposite to that in the concertina, requiring:

$$R_{L(\text{input stage})} = \frac{(\mu + 2) + r_a}{\mu} \cdot R_{L(\text{concertina})}$$

The previous equation is simply $R_{L(\text{concertina})}$ divided by the gain of the concertina to one output, thus ensuring equal currents in the load resistances of each stage. A combination of $R_L = 33\,k\Omega$ for the input stage and

$R_L = 27\,k\Omega$ for the concertina allows this, and results in $I_a = 6\,mA$ for the input stage which reduces noise. In this respect, the amplifier is technically flawed, and should be completely rebuilt, but the author enjoys listening to music through it, and doesn't feel inclined to fiddle.

Now that we know the HT voltage for the two stages, we can calculate all the AC parameters.

Calculation of the cathode bias resistor and feedback resistor

This is easily the most complex calculation in the design of a power amplifier with negative feedback applied to the cathode of the input stage. However, if we stay calm, and draw lots of neat diagrams, *fully labelled* with every piece of information that we have, then the problem is reduced to manageable proportions.

Attempting to solve this problem on the back of a small envelope is doomed to failure.

We have four main factors at work:

- We need to set our cathode bias voltage correctly. This would normally be a trivial application of Ohm's law, but in this case the bias current flows through the cathode resistor *and* the feedback resistor.
- The input valve itself generates a feedback current through the cathode resistor in addition to any current sourced from the output of the amplifier.
- We need to set the ratio of the two resistors so as to obtain the desired amount of negative feedback.
- As far as AC is concerned, the cathode resistor is shunted by r_k of the valve.

Now that we have stated our restrictions, we should be able to label a diagram and derive some equations.

Since we want 2.5 V bias on the cathode, and anode current is 190 V/47 kΩ, the total resistance to ground from the cathode must be 618.4Ω.

The anode signal swing required for the full claimed Mullard output of 11 W is $8.636\,V_{RMS}$. This means that the anode signal *current* must be $8.636\,V/47\,k\Omega = 0.1837\,mA_{RMS}$. This current also flows in the cathode circuit and will develop a feedback voltage across any unbypassed cathode resistor.

We wish to make the input sensitivity of the amplifier $2\,V_{RMS}$, and we know that the unmodified sensitivity is $298\,mV_{RMS}$, so the feedback voltage required

at the cathode will be $2 - 0.298 = 1.702\,V_{RMS}$. We know that for the full output of 10 W, the signal at the output of the amplifier will be $8.944\,V_{RMS}$. Therefore there will be $7.242\,V_{RMS}$ across the feedback resistor.

Since r_k will shunt the cathode resistor at AC, we need to find r_k:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

Using this equation, we find that $r_k = 1.559\,k\Omega$.

We will assume that the output of the amplifier is a true Thévenin source driving the network through the feedback resistor '*y*'. We will represent the valve's own feedback current as a Norton current source, and we will shunt the cathode resistor '*x*' with r_k. We will label our currents and include *every* piece of information available. Note that this is an AC diagram. See Fig. 6.34.

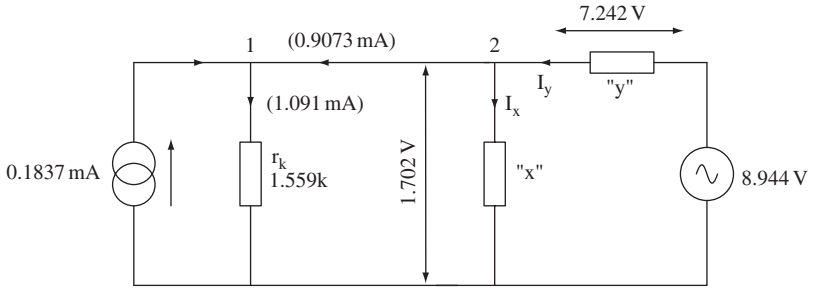


Fig. 6.34 *Equivalent circuit of AC conditions at the input stage*

From now on, the solution is easy, if a little tedious. Our first observation is that there is a resistor of known value (r_k) with a known voltage (1.702 V) across it, so we ought to calculate the current through it. This current is 1.091 mA, and we can label it on the diagram.

Having labelled this current, we can now see that node 1 has two known currents flowing through it, so we can find the third, using Kirchhoff. If there is 0.1837 mA flowing into the node, but 1.091 mA leaving it, then 0.9073 mA must be supplied by the other node.

Moving to node 2, we can see that any current coming into the node must be supplied by I_y, and that this splits through the resistor '*x*', and to node 1. Formalizing this:

$$I_x + 0.9073 = I_y$$

We can use Ohm's law to make statements about the currents in resistors 'x' and 'y':

$$I_y = \frac{7.242}{y}$$

$$I_x = \frac{1.702}{x}$$

The final restriction is the DC restriction, which says that x and y in parallel $= 0.6184\,\text{k}\Omega$.

$$0.6184 = \frac{xy}{x+y}$$

Now that we have some equations, we can solve them. This is easy, generating the equations was the hard bit! The way to do this is to substitute the second and third equations into the first:

$$\frac{1.702}{x} + 0.9073 = \frac{7.242}{y}$$

Rearranging, and simplifying:

$$7.982x - 1.876y = xy$$

We can now substitute this into the fourth equation, and solve it to give the ratio $y = 2.953x$. Substituting this ratio back into the equation yields $x = 828\,\Omega$, and using the ratio, we find that $y = 2.44\,\text{k}\Omega$.

Using standard values, we therefore need $1.2\,\text{k}\Omega$ in parallel with $2.7\,\text{k}\Omega$ for the cathode resistor, and $4.7\,\text{k}\Omega$ in parallel with $5.1\,\text{k}\Omega$ for the feedback resistor.

Some of V_1's cathode current flows through the output transformer, and it might be thought that this would cause distortion. Assuming that the DC resistance of the transformer secondary winding is negligible compared to the feedback resistor, the current flow will be $2.5\,\text{V}/2.44\,\text{k}\Omega \approx 1\,\text{mA}$. The current turns ratio of the transformer is 31.6:1 (secondary to primary), so 1 mA of DC flowing in the secondary is equivalent to $31\mu\text{A}$ out of balance DC flowing in the primary. Compared to 40 mA each side, this is entirely negligible, since output valve balance is highly unlikely to be as good as this.

We now know all the component values for the driving circuitry, so we need to determine values for the output stage.

The EL84 is allowed a maximum grid leak resistor of 300 kΩ with grid bias, but as we are using cathode bias, we can cheerfully increase this to 470 kΩ. We then need a 0.1 μF coupling capacitor, which should be polycarbonate, or preferably polypropylene, with a rating of $\geq 400\,V_{DC}$.

4.7 kΩ is the traditional value for grid stopper resistors on the EL84; it is possible that we may not actually need them, but it seems sensible to fit them just in case.

47 Ω in series with g_2 is alleged to reduce distortion whilst reducing peak power, the author has not tested this, so fitting them is a matter of personal choice. The Mullard circuits did fit them, whilst the Leaks didn't.

The cathode bias resistor should be 270 Ω (from the Mullard data sheet), and this will dissipate 0.45 W. 2 W resistors are commonly used here, but a 15 W chassis mounting aluminium-clad type with tabs is a better practical choice. This is because we will place an electrolytic capacitor very close to the resistor, so it must be kept cool. Additionally, the chassis-mounting resistor provides convenient tags for anchoring the capacitor.

The cathode bypass capacitor should be 2200 μF for a 1 Hz cut-off, but as discussed earlier, this value would exacerbate recovery problems after overload, so we will compromise on 470 μF 63 V. 63 V may seem excessive, since it will only see ≈11 V, but the higher voltage component has a lower ESR, which becomes significant when we are trying to bypass the 67.5 Ω seen at the valve ($r_k = 90\,\Omega$, $R_k = 270\,\Omega$).

Because we have only one RC network plus the output transformer in the entire amplifier, LF stability will not be a problem. HF stability is not assured, so this should be investigated.

The input stage has its basic sensitivity reduced from 298 mV to 2 V, which corresponds to a gain reduction of 6.71 or 17 dB. From this, we can calculate the new r_a for the stage:

$$\frac{\mu R_L}{R_L + r_a} = 6.71\frac{\mu R_L}{R_L + R'_a}$$

Solving this, and using $r_a = 5\,k\Omega$, gives $r'_a = 302\,k\Omega$, in parallel with $R_L = 47\,k\Omega$, this gives $Z_{out} = 41\,k\Omega$. Applying global negative feedback invariably causes $Z_{out} \approx R_L$ for the input stage.

The concertina has 3.2 pF of Miller capacitance, and if we allow for strays, 5 pF is a reasonable total value. In combination with 41 kΩ, this gives a cut-off of about 780 kHz.

The output stage has input capacitance that loads the output resistance of the concertina. Although the EL84 is a pentode, it still has Miller capacitance, albeit greatly reduced, so this should be included in the calculation.

We can find the anode gain of the output stage by calculating the voltage across the 8 kΩ transformer primary for 11 W, and knowing that we need 16 V_{RMS} from grid to grid to drive the stage. This gives a gain to the anode of 18.54. Since $C_{ag} = 0.5$ pF, this would result in a Miller capacitance of 9.8 pF. Unfortunately, this value of C_{ag} is for the pure pentode connection, whereas we are using the Blumlein or 'ultralinear' connection, where g_2 does *not* stay at a constant potential. We should therefore make an allowance for Miller effect from C_{g2}, unfortunately, the Mullard data sheet does not give a value for this, but it is probably wise to allow another 10 pF. If we add these to $C_{in} = 10.8$ pF, we have a total input capacitance of ≈35 pF including strays. Provided that the output stage never strays out of Class A1, this combines with concertina output resistance (≈190 Ω) to give a cut-off frequency of 24 MHz. We can easily afford to slug this slightly since any additional capacitance will swamp the variations in capacitance between valves, and will therefore improve HF balance. Shunt capacitors of 68 pF across the EL84 grid leak resistors slug this pole to 8 MHz, and tend to preserve equal loading on the concertina if the output stage strays into Class B.

We can now draw a full circuit diagram of the amplifier, with component values. See Fig. 6.35.

The author's prototypes

The author has found this design of EL84 amplifier to be very satisfactory, and has built five variations on the circuit. A pair of scrapped Leak TL12+ contributed the output transformers for the No. 2 amplifier, whilst a pair of surplus mains transformers and chokes were used for the HT, hence the GZ34 rectifiers. See Fig. 6.36.

For the last nine years, this amplifier has powered a pair of Rogers LS3/5a and is driven directly by a Marantz CD63 mkII KI. A later version drives tweeters in the author's main system. All but the original amplifier used output transformers from Leak Stereo 20 or TL12+, and Nos 3–5 used the mains transformer and chassis as well, but the author has experimented with component types.

- No. 2: ideal for LS3/5a: produces a traditional rounded valve sound. Metal film anode load resistors, polycarbonate coupling capacitors, but polypropylene HT capacitors.
- No. 3: a beautifully sweet 5 W: metal film load resistors, polycarbonate coupling capacitors, electrolytic HT capacitors, with a 15 H choke

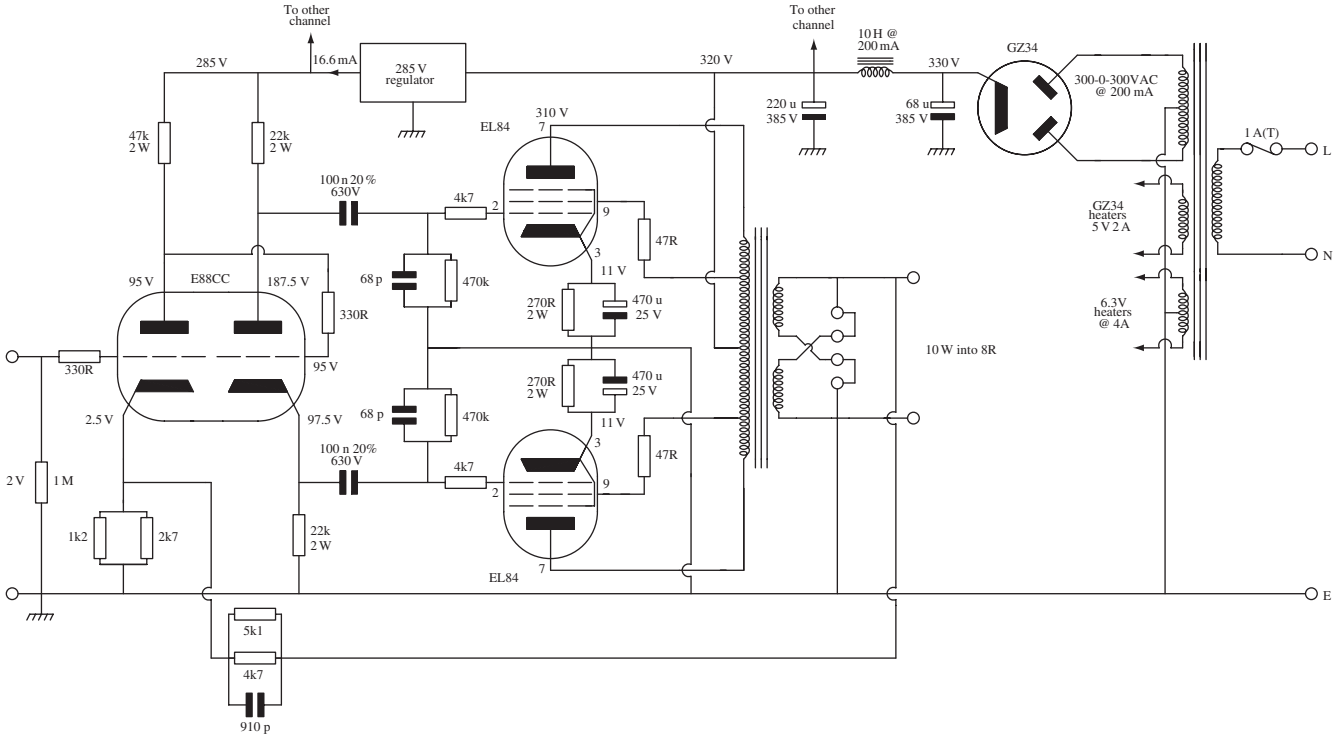


Fig. 6.35 *Practical 'Bevois Valley' power amplifier*

Fig. 6.36 *The author's prototype*

replacing Leak's 100 Ω resistor between reservoir and load capacitor. Output valves strapped as triodes ($R_k = 560\,\Omega$).

- No. 4: a more modern, analytical, but less forgiving sound: aluminium cased wirewound anode load resistors, polypropylene coupling capacitors, and polypropylene HT capacitors.
- No. 5: aluminium cased wirewound anode load resistors, polypropylene coupling capacitors, polypropylene HT capacitors, 5 H choke replacing Leak's resistor, ceramic valve bases.

The traditional approach to power >10 W

The traditional method of increasing power was to use a more powerful valve such as the Mullard EL34, or even the GEC KT88. Another method is to use Class AB; using these techniques, we can obtain 50 W from a pair of Mullard EL34 or GEC KT66, and 100 W from a pair of GEC KT88. After this, we resort to transmitter valves at enormous cost.

This traditional method has a number of disadvantages:

- Transmitter power valves are invariably disproportionately more expensive.
- They use high HT voltages – so the smoothing capacitors are expensive, and the HT supply is a major safety hazard.

- They tend to need higher impedance anode loads – making the design of a good transformer more difficult.
- They have savage drive requirements – often needing a power valve as a driver.

However, there are ways of avoiding these problems.

Sex, lies and output power

In the late 1960s and early 1970s, some quite unpleasant audio amplifiers were made using *transistors*. Compared to the valve behemoths, these transistor amplifiers were very small and light, but they didn't actually sound any better (in fact, most sounded a lot worse), so something was needed to make them sell. The one thing that early transistor amplifiers could do was to provide plenty of power, and thus the power rating war started.

To make a truly powerful amplifier, a large power supply is needed, but this is expensive. Now (classical) music generally only has short duration peaks, and nobody listened to anything else (or at least nobody whose opinions were taken seriously), so amplifiers were designed that could manage higher output powers, but only for a very short time. This allowed power ratings to be increased further, and the 'music power' rating was born. We measure the maximum output power at 10% distortion, or the onset of *clipping* (the point at which a sine wave begins to have its peaks clipped off), with bursts of 1 kHz driving one channel *only* into a resistive load. By this means, it is perfectly possible to convert a 20 W amplifier with a poor power supply into a 50 W model, and if we now double the output to account for two channels, we have a 100 W amplifier.

At least four fallacies were used in the previous argument, but they were nothing compared to the outrageous power claims made for many computer loudspeaker systems. A recent example having a woofer the size of a large loaf of bread and a pair of small loaf satellites claimed a power rating of 800 W PMPO and all for £23! (PMPO = Peak Music Power Output, or in this case, Purely Mythical Power Output.)

Loudspeaker efficiency and power compression

We *can* build more efficient loudspeakers. This is an excellent solution, since inefficient loudspeakers frequently suffer from power compression; an effect whereby the resistance of the voice coil rises due to temperature, and reduces sensitivity until the coil has cooled down.

Active crossovers and Zobel networks

We can drive the loudspeakers more effectively. If the drive units are driven by dedicated amplifiers preceded by an active crossover, many benefits result.[12] For the purposes of this discussion, it is sufficient to say that a two-way loudspeaker system, driven actively by 10 W amplifiers will go surprisingly loud, very cleanly. (Details of active crossover design are given in the Appendix.)

However, there can occasionally be a problem. Modern moving coil tweeters tend to be Ferrofluid® damped, resulting in a very nearly resistive electrical impedance, but bass drivers cannot often use this technique because their greater cone/dust cap excursion sufficiently compresses the air within the magnet assembly to explode any Ferrofluid® out of the gap.

Bass loudspeakers have voice coils with significant self-inductance, so their dedicated amplifier sees a rising impedance which can compromise HF stability. Additionally, beam tetrodes and pentodes produce higher amplitude higher harmonics in their distortion spectrum as load resistance rises – so correcting voice coil inductance to the optimum impedance load would be worthwhile. Fortunately, a simple moving coil loudspeaker is easily corrected by adding a Zobel network directly across its terminals. See Fig. 6.37.

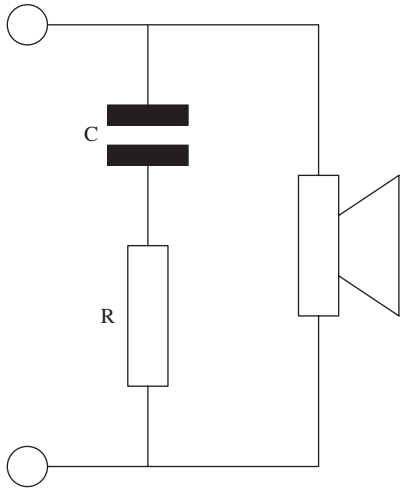


Fig. 6.37 *Zobel network for cancelling voice coil inductance*

The required additional resistor is equal to the DC resistance of the loudspeaker, and the capacitor value is found using:

$$C_{Zobel} = \frac{L_{voice\ coil}}{R_{DC}^2}$$

Because the loudspeaker can be considered to be a transformer with its voice coil loosely coupled to the shorted turn of the pole pieces which have hysteresis losses, the simple model of pure inductance in series with resistance is somewhat inaccurate, but sufficiently good for our present purpose.

Parallel output valves and transformer design

This is a cracking solution, and gives many advantages. If we use multiple pairs of paralleled output valves, we can keep the HT voltage within reasonably safe bounds, perhaps even at 320 V, if we are prepared to use many pairs of EL84. With each additional pair of valves, the transformer primary impedance falls, as does the turns ratio, making it easier to design a good quality component. Statistically, total anode current per side will be better balanced as we increase the number of valves, and deliberate selection will improve this still further.

Driving higher power output stages

Whether they are composed of paralleled devices or not, higher powered output stages always require more of the driver circuitry. When we investigated the Williamson amplifier, we found that it had a dedicated driver stage, but that the large number of stages made stability a problem. Clearly, a better approach is needed.

As before, listing the requirements helps solve the problem.

- We need low output resistance to drive the increased input capacitance of the output valves, and a cathode follower may be needed.
- We need to provide a large output voltage with low distortion, this invariably demands some form of a differential pair.
- Wide bandwidth and high gain are also desirable, because we would like to have only one set of coupling capacitors to ensure LF stability, and the cascode might be ideal, although a carefully designed cascade of DC coupled differential pairs could be even better.

We will first investigate a cascode differential pair with direct coupled cathode followers, sometimes known as the Hedge[13] circuit, after its designer

(although the original Hedge circuit did not include cathode followers). See Fig. 6.38.

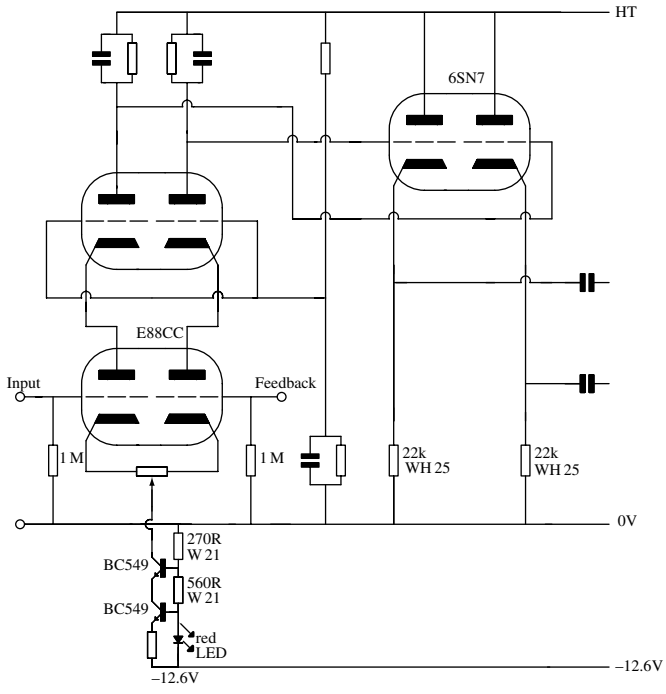


Fig. 6.38 *'Hedge' cascode differential pair plus direct coupled cathode followers*

Design of the individual parts of this circuit was covered in Chapter 2, so we need not go into great detail on this circuit other than to make a few observations.

A single differential pair is not the ideal phase splitter, so we must take extra care over this to obtain a good result. The anode load resistors must be matched, aged and generously rated to avoid drift. The constant current sink should be made to have as high an output resistance as possible, and stray capacitance to ground from the cathode should be minimized to maintain a high impedance at HF. Matching the valves would be useful if possible.

Each pair of valves requires a separate heater supply. Sad, but true. The cathode followers need $\approx$200 V superimposed on their heaters, the upper pair of the cascode need $\approx$100 V, and the lower pair 0 V. Flirting with this rule will generate problems related to heater cathode insulation breakdown/

leakage, and emission from the heater to the cathode will be summed with the intentional cathode current. You have been warned!

As was mentioned before, the only really satisfactory valve for use as the lower valve in a cascode is the E88CC, any other type will waste HT. The cathode voltage on the lower valves is usually quite low, $\approx 2.5\,V$, and because phase splitters inevitably have half the input signal voltage on the cathode, the tail of the sink needs to be taken to a subsidiary negative supply.

Feedback from the output can be applied to a grid, which makes the calculations of feedback network much easier, or the stage could accept a balanced input.

The 'Crystal Palace' amplifier

As with the 'Scrapbox Challenge' and 'Bevois Valley' amplifiers, the design of a larger push–pull amplifier begins with the output valve. Once all of the soft-start and general safety considerations have been taken into account, the cost of an HT supply is proportional to the square of its voltage. Thus, reducing the HT voltage releases money that can be spent elsewhere to achieve a better overall set of compromises. Possible contenders for the output stage are:

	845	**813**	**4 × EL34**	**13E1**
$P_{a(max.)}$	100 W	100 W	100 W	90 W
$P_{g2(max.)}$	–	22 W	32 W	10 W
$I_{k(max.)}$	120 mA	180 mA	600 mA	800 mA
$V_{a(max.)}$	1250 V	2250 V	800 V	800 V
$V_{g2(max.)}$	–	1100 V	500 V	300 V
μ	5.3*	8.5*	10.5*	4.5*
gm	3.4 mA/V*	4 mA/V†	46 mA/V*	35 mA/V†
r_a	1.6 kΩ†	2.1 kΩ†	230 Ω*	130 Ω*
V_h	10 V	10 V	6.3 V	26 V
P_h	32.5 W	50 W	37.8 W	33.8 W
C_{ag}	12.1 pF*	17 pF‡	44 pF‡	40 pF‡
C_{miller}	76 pF	162 pF	500 pF	220 pF

* = manufacturer's claimed value.
† = value derived from manufacturer's data sheet.
‡ = value measured by author.
Note that the values in this table apply to NOS valves, and may not be applicable to recently manufactured valves.

Of these valves, the 845 is a true triode, the 813 is a triode-strapped beam tetrode, the quartet of EL34 is a triode-strapped pentode, and the triode-strapped beam tetrode 13E1 actually contains a duet of paralleled valves.

All the options offer $P_{a(max.)} \approx 100\,W$, so they could all achieve approximately the same output power. NOS 845 are extremely expensive, but the modern version is only rated at $P_{a(max.)} = 75\,W$. NOS 813 are readily available, but require the same expensive HT as the 845 ($\approx 1000\,V$). Sadly, the 13E1 is more expensive than a quartet of EL34. Nevertheless, when the author saw a 13E1, it was lust at first sight. You will probably be more rational, and opt for a quartet of EL34.

13E1 conditions

Push–pull output stages can be analysed using *composite* curves.[14] Composite curves are obtained by placing a second set of curves back to back below the first. Fictitious lines are then drawn between opposite true anode curves, and these are deemed to be the composite anode curves. See Fig. 6.39.

The operating point is where the composite line from $V_{gk} = -60$ V of V_1 to $V_{gk} = -60$ V of V_2 passes through $I_a = 0$ at $V_a = 250\,V$. For maximum power output, $R_L = 2r_a$, and this loadline can be drawn by mirroring the composite $V_{gk} = -60$ V line about a vertical line passing through the operating point. In this particular instance, $R_L = 277\,\Omega$, and the predicted power output is 42 W. Note that this particular operating point implies Class AB operation, since $V_a = 250\,V$, $V_{gk} = -60\,V$ implies $I_a = 49\,mA$. Quite apart from any reservations we might have about Class AB, the extremely steep loadline produced by this method greatly increases odd harmonic distortion. Although composite curves are a useful theoretical concept, and demonstrate the difference between Class A and Class B very clearly, they imply ideal valves and are quite fiddly to produce and adjust, even on a computer.

We will find that it is much easier to analyse one half of the output stage, treating it as a single-ended stage. Once we have discovered the optimum single-ended loadline, we simply convert that into the push–pull requirement. In theory, we lose some accuracy by not using composite curves, but precise loadlines in power stages are futile because loudspeakers are not pure resistances, so the saving in drawing effort is well worthwhile.

Because this is an output stage, and we want to extract maximum power, we must operate the valve at $P_{a(max.)} = 95\,W$. In general, when shuffling

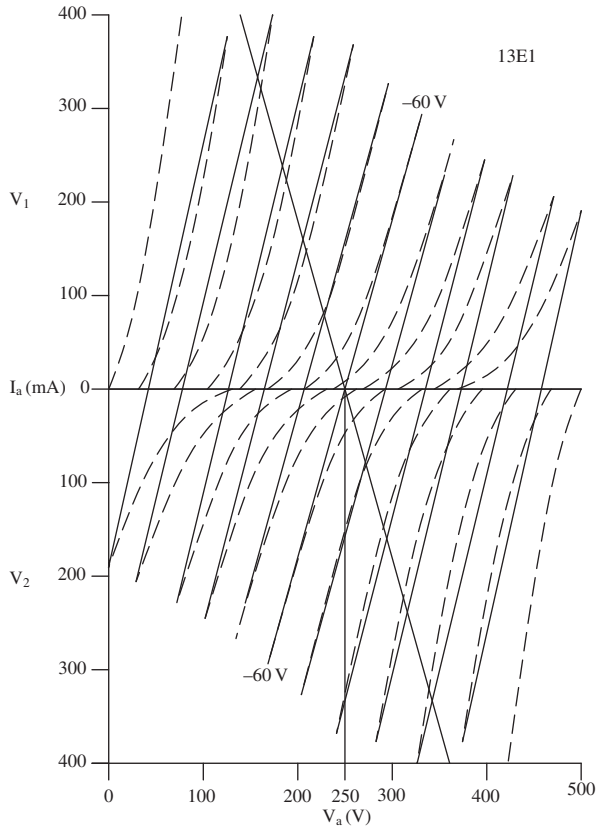


Fig. 6.39 *Composite anode curves for push–pull stage*

loadlines and operating points for a given valve, we will find that output power is proportional to anode voltage, whilst distortion is inversely proportional. But cost rises dramatically with anode voltage, so we will set $V_a = 400\,V$. Since $P_{a(max.)} = 95\,W$, and $V_a = 400\,V$, we can use $P = IV$ to calculate that $I_a = 237.5\,mA$, and plot this point. See Fig. 6.40.

There is an anode curve near to our operating point, so we can find r_a. In this instance, we find that $r_a = 282\,\Omega$. (We should not worry that this is significantly poorer than the manufacturer's claimed value – they typically measure at $V_{gk} = 0$, $I_a = I_{a(max.)}$.) Traditionally, we set $R_L = 2r_a$ for maximum power, so the author tried this loadline. After extrapolating the curves (plausibly making them up), the 564 Ω loadline offered 14 W from the valve. Since $P_a = 95\,W$, this didn't seem too promising, but a gentler loadline of 625 Ω predicted 20 W with lower distortion at the same dissipation.

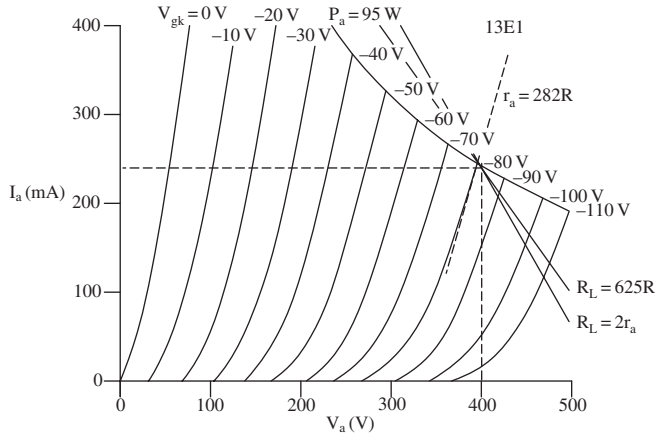


Fig. 6.40 *Setting the 13E1 operating point*

In the push–pull amplifier, each output valve must see a load of 625 Ω, so each half of the transformer is wound with the correct number of turns to reflect this load. However, push–pull operation halves anode resistance. Since reflected impedances change by the square of the turns ratio, doubling the number of turns quadruples the impedance. Thus, our output transformer would measure $4 \times 312\,\Omega = 1.25\,\text{k}\Omega$ from anode to anode. If each valve contributes 20 W, the total output power must be 40 W.

Eagle-eyed readers will note that $V_{g_2(\text{max.})} = 300\,\text{V}$, but that the final design sets $V_{g_2} = 400\,\text{V}$. Strapping pentodes and tetrodes as triodes and then exceeding their $V_{g_{2(\text{max.})}}$ rating has been done before, most notably by Langford-Smith, using a pair of triode-strapped 807 at 400 V ($V_{g_2(\text{max.})} = 300\,\text{V}$) to replace KT66 in a Williamson amplifier. More significantly, Philips[15] gave performance data for a triode-strapped QE 05/40 beam tetrode audio amplifier operating at 400 V despite the fact that they quoted $V_{g_2(\text{max.})} = 250\,\text{V}$ for the same valve used as a beam tetrode in an audio amplifier.

Nevertheless, the author still has some misgivings about exceeding $V_{g_{2(\text{max.})}}$ ratings, so the emphasis in this design is on designing driver circuitry of irreproachable performance that could drive any of the valves listed in the table.

Driver requirements

The stipulation 'irreproachable performance' is very vague, and needs to be converted into engineering requirements that will lead to engineering solutions:

1 Minimal measured distortion.
2 Distortion to be composed of low order harmonics.
3 Push–pull output with good balance.
4 Large undistorted voltage swing.
5 Sufficient gain to allow global negative feedback if required.
6 Low DC output resistance to avoid problems with DC grid current.
7 Low AC output resistance to drive load capacitance.
8 Tolerance of output stage conduction angle changes from 360° to 0°.
9 Instantaneous recovery even after gross overload.

Finding a topology that satisfies the driver requirements

1 Minimal measured distortion

This requirement implies nearly horizontal loadlines. A horizontal loadline implies an active load, but large resistive loads requiring a higher HT voltage are also a possibility.

2 Distortion to be composed of low order harmonics

This requirement implies triodes rather than pentodes. Taking requirements (1) and (2) into account simultaneously suggests that triodes from the *SN7/*N7 family would be ideal.

3 Push–pull output with good balance

This requirement is best solved by two cascaded differential pairs with constant current sink tails. Since triodes produce predominately 2nd harmonic distortion, which the differential pair cancels, this satisfies requirement (1), but reinforces the preference for the *SN7/*N7 family because valves producing low 3rd harmonic distortion are now needed due to odd harmonics summing constructively in a differential pair.

4 Large undistorted voltage swing

One of the strengths of the differential pair is its linearity when swinging large voltages. Nevertheless, the more HT voltage available, the better, so this requirement implies that the driver stage should have HT $> 400\,V$. Since the output stage is likely to use $\approx 400\,V$, this implies that the driver stage should have a dedicated HT supply.

5 Sufficient gain to allow global negative feedback if required

This requirement can probably be satisfied by two cascaded *SN7/*N7 differential pairs. If necessary, the gain could easily be doubled by using a high μ valve dual triode such as 6SL7, 7F7, ECC83, or ECC808 in the input differential pair.

6 Low DC output resistance to avoid problems with DC grid current

Most of the larger power valves pass significant grid current even when the grid is negative, which is why manufacturers' data sheets recommend such low maximum values for grid leak resistors. But a small grid leak resistor is an unnecessarily harsh load for the preceding stage.

Satisfying this requirement demands that the drivers must be DC coupled to the output stage grids. Output stage HT is used most efficiently if the cathodes of the output valves are at 0 V because this means that $V_{HT} \approx V_a$. Therefore, V_a of the driver stage must be negative to bias the output valves correctly. The anodes of the driver differential pair can only be at a negative voltage if the tail of the differential pair is returned to a substantial negative HT supply, perhaps −300 V. If the driver stage uses the output stage supply for its positive HT, it now has a rail-to-rail HT of 700 V, which easily satisfies requirements (4) and (1).

7 Low AC output resistance to drive load capacitance

Although the *SN7/*N7 family produces low distortion, r_a is not particularly low, and fails this requirement. Valves such as the 6BX7 and 6BL7 have lower r_a, but their distortion is high, and their Miller capacitance punitive. Adding cathode followers to the outputs of the differential pair divorces the responsibilities for low distortion and low output resistance, allowing the differential pair to be optimized for linearity and swing, and the cathode followers for current driving ability.

8 Tolerance of output stage conduction angle changes from 360° to 0°

There is more to meeting this requirement than at first seems evident. When we investigated phase splitters, we found that all phase splitters were sensitive to their loading, requiring Class A loads. The requirement for cathode followers has now been reinforced, since their buffering action allows the two-stage phase splitter to operate undisturbed by arbitrary output stage conduction angles.

The author considers that attempting to drive an output stage cleanly into Class AB2 is not worth the candle, so the cathode followers will be biased only so that they can drive the output stage Miller capacitance cleanly, and no

attempt to drive grid current will be made. To maximize output swing, the cathode followers are likely to be operated with their cathodes at half the rail-to-rail HT voltage, so $V_a = 350\,V$. If we do not attempt to drive grid current, $I_a = 7\,mA$ is adequate, resulting in $P_a = 2.5\,W$, which is just within the range of the *SN7/*N7 family (see Appendix).

9 Instantaneous recovery even after gross overload

This requirement means that the amplifier must not suffer from blocking. Therefore, the coupling capacitors must be positioned so that they couple to a stage that cannot be overloaded. By definition, the output stage can be overloaded, but we have already specified that it must be DC coupled. There is no advantage to placing the coupling capacitors between the second differential pair and cathode followers, because the anodes of the differential pair need to be at roughly the same voltage as the grids of the cathode followers in order to take advantage of the rail-to-rail HT voltage. The ideal position for the coupling capacitors is *between* the two differential pairs.

An early iteration of the driver circuitry achieved 0.03% THD + N just below the point where output stage grid current imposed catastrophic loading. The author is simultaneously embarrassed and proud to report that measuring its distortion whilst it cleanly swung a differential voltage of +47 dBu ($177\,V_{RMS}$) into a $100\,k\Omega$ load briefly gave a figure of 0.11% THD + N before the MJS401D briefly flashed 'LEVEL HIGH' and died. Moral: don't get carried away and abuse your test equipment.

Even though each proposed 13E1 requires $\approx 58\,V_{RMS}$, the driver stage has $\approx 10\,dB$ of headroom at *all* frequencies. As a consequence, the output stage can be overloaded by 10 dB (causing gross distortion) before the second differential pair suffers blocking, so recovery from any reasonable overload *is* instantaneous.

Summarizing the outcomes of the requirements, we need a cascade of differential pairs separated by coupling capacitors, using valves from the *SN7/*N7 family, and powered by a split-rail HT supply. The output of the driver differential pair will be DC coupled to cathode followers, which will be DC coupled to the output stage grids. See Fig. 6.41.

Circuit topology: power supplies and their effect on constant current sinks

The *SN7/*N7 family of valves produces primarily 2nd harmonic distortion, which can be cancelled in a differential pair *if* no signal current is lost in the

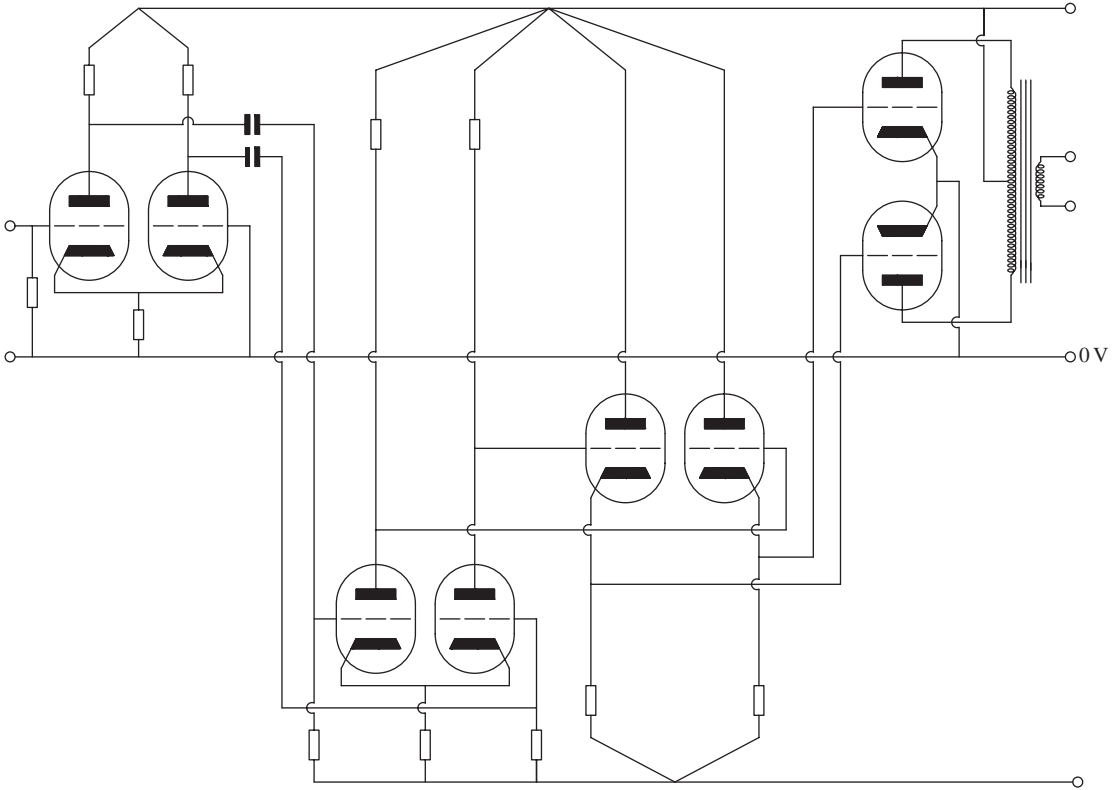


Fig. 6.41 *Conceptual diagram to show topology and position of coupling capacitors*

tail. Thus, we need active tails for both differential pairs, but because the grids are capacitor coupled from the previous stage, their grids are returned to the same supply as the constant current sink, V_k must be low, so these must be semiconductor constant current sinks.

The second differential pair is likely to have $\geq$500 V of HT voltage, so V_{gk} is likely to be ≈ -10 V, which allows sufficient voltage for a cascode constant current sink to operate without an additional supply. Sadly, the first stage is likely to have quite a low HT voltage, so there is insufficient V_{gk} for a cascode, but a 334Z constant current sink is ideal. However, the 334Z has an absolute maximum current rating of 10 mA, whereas we can design our cascode to pass any current we like. Thus, our choice of valve operating points has already been restricted.

$V_{a(max.)}$ and the positive HT supply

We have elected to direct couple the cathode followers to the grids of the output valves, so their cathodes will be at ≈ -82 V, depending on individual output valves. If the anodes of the cathode followers are connected to the output stage HT, $V_{ak} = 482$ V, which we cannot allow. (Even for the GTA or GTB versions of the *SN7, $V_{a(max.)} = 450$ V.) However, this problem is not as bad as it sounds, because we do not need the cathode follower to swing 482 V_{pk}, so we can lower our positive supply to 160 V, which reduces V_{ak} to $\approx$250 V, allowing any valve from the *SN7/*N7 family to be used.

Next, we must consider the DC biasing of the output stage. The high *gm* of our chosen output stage (whether 13E1, or a quartet of EL34) means that the output stage current is extremely sensitive to changes in V_{gk}, and 30 mA/V is a very high mutual conductance in valve terms, so we cannot permit the grid bias voltage to drift. Because the driver circuitry is DC coupled from the output valves to the anodes of the second differential pair, a change in its V_a could potentially damage the output valves since (by definition) they operate at $\approx P_{a(max.)}$.

AC considerations dictate that the differential pairs require constant current sink tails. Once the driver differential pair has anode loads made of (constant value) resistors, Ohm's law ensures that unchanging V_a can be achieved by regulating the 160 V positive HT supply. Our design has now evolved to the point where it *requires* an HT regulator to work safely. Drawn in full on a circuit diagram, regulators are intimidating, but only a valve regulator is more expensive than a decent HT capacitor and series resistor. In this instance, a valve regulator is probably not suitable (excessive long-term DC drift due to steps in neon reference voltage), so we should not be worried by needing a semiconductor HT regulator.

Symmetry and the negative HT supply

Because cathode followers operate under 100% negative feedback, they contribute very little distortion compared to the second differential pair, but because they drop $\approx 8\,V$ across V_{gk}, they modify the anode voltage of the second differential pair from $-82\,V$ to $-90\,V$. Since we require maximum linear swing from the second differential pair, the negative HT should be symmetrically opposite its anode voltage, requiring a negative HT of $-90\,V - 260\,V = -350\,V$. The negative HT voltage is not at all critical and does not even need to be regulated because variations simply change V_{ak} without affecting I_a.

Although the precise voltage of the negative HT is not critical, it is essential that this supply is reliable. Failure would drive the output valve grids positive and the resulting anode dissipation would quickly destroy them. Thus, not needing a regulator for the negative HT has the bonus of improved reliability. Nevertheless, the output stage includes an HT fuse for protection in the event of negative HT failure.

The second differential pair and output stage current

Now that we have some firm HT voltages, we can begin detailed audio design, working back from the cathode followers towards the input stage.

The *SN7/*N7 family offers optimum linearity when $I_a \geq 8\,mA$. The voltage across R_L for the cathode follower is $-82\,V - -350\,V = 268\,V$, so Ohm's law dictates that a $33\,k\Omega$ 6 W wirewound resistor dissipating 2.2 W would achieve $I_a = 8.1\,mA$.

Furthermore, the 13E1 presents 220 pF of Miller capacitance which must be driven cleanly. At 20 kHz, the reactance of 220 pF is $36\,k\Omega$. At full power, the output stage demands $58\,V_{RMS}$, and maintaining this voltage across the reactance of the 220 pF capacitance demands $1.6\,mA_{RMS}$, or $2.3\,mA_{pk}$. The capacitive load forces anode current to swing vertically $\pm 2.3\,mA$ on the loadline, which requires *gm* to be constant. Fortunately, at $I_a = 8.2\,mA$, *gm is* reasonably constant.

$V_a = 160\,V - -82\,V = 242\,V$, and we already knew I_a, so we can determine V_{gk} at these conditions. Referring to the curves, $V_a = 242\,V$ and $I_a = 8.1\,mA$ roughly intersects the $V_{gk} = -8\,V$ curve. Knowing this voltage is important because it means that we now know that the grid voltage is $-82\,V- \approx 8\,V \approx -90\,V$. The grids are DC coupled from the anodes of the second differential pair, so their required anode voltage is also $-90\,V$.

For the second differential pair, the voltage across R_L is $160\,V - -90\,V = 250\,V$. Each triode in the second differential pair can

pass <8 mA because most of its distortion will be cancelled by push–pull action (this is not true for the cathode followers when the output stage enters Class AB). 50 kΩ MPC-5 anode load resistors are convenient, so the current required to set V_a to −90 V is 250 V/50 kΩ = 5 mA, so the total tail current is twice this at 10 mA.

The tail current is highly significant. Increased tail current causes increased voltage drop across the anode load resistors of the second differential pair, causing their absolute voltage to become more negative. The cathode followers faithfully follow this negative change, and so the grids of the output valves become more negative, reducing their anode current. Thus, making the tail current adjustable allows us to set output stage current.

The output valves may not be perfectly matched, so interposing a variable resistor between the cathodes of the second differential pair allows us to adjust the balance of the stage, and therefore output stage current balance. See Fig. 6.42.

Why not have tighter stabilization?

Since we saw earlier that the output stage was sensitive to changes in V_{gk}, and that this is set by tail current, it seems intuitively obvious to stabilize tail current as tightly as possible. However, we should consider the effect of an increase in mains voltage in more detail. When mains voltage rises, the negative HT rail becomes slightly more negative, and more current flows through the resistor chain supplying the reference LED, so its voltage across it slope resistance rises slightly. But V_{be} for the transistor is unchanged, so the voltage across the current programming resistor rises identically, and tail current increases. Increased tail current *reduces* output stage current, but the rise in mains voltage caused the (unregulated) HT to the output stage to rise, which would have *increased* current, so the two effects oppose on another, which is desirable. Thus, we discover that tightly stabilizing the tail current and negative HT would be counter-productive because it would also require stabilization of the high current HT supply to the output stage.

As mains voltage varies, not only does it change the HT voltage, but also the heater voltage. The DC conditions of the differential pairs are forced to be correct by the constant current sources and HT regulation, and the cathode followers have plenty of feedback, but the output valves are sensitive to heater voltage. Fortunately, because the 13E1 heaters can be configured to operate from 26 V, a pair of valves requires only 2.6 A, which can be regulated reasonably efficiently.

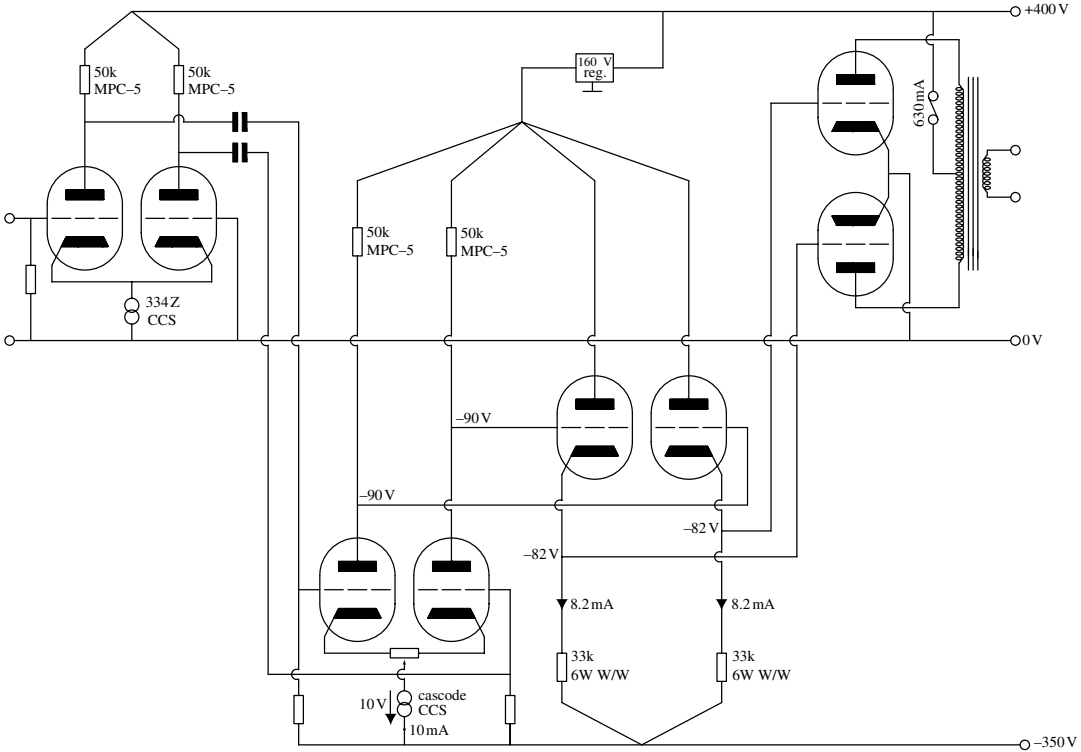


Fig. 6.42 *Setting the DC conditions for the amplifier*

The first differential pair, its HT supply and linearity

By comparison with the second differential pair, the considerations involved in the design of the differential pair of the input stage are trivial. The stage only has to furnish $\approx 3.3\,V_{RMS}$ from each output, so distortion really isn't a problem. Nevertheless, the *SN7/*N7 family requires $V_a \geq 150\,V$ for reasonable linearity even at very small voltage swings, so a $>300\,V$ HT would be ideal.

The second differential pair required a negative HT of $-350\,V$, and if a traditional centre-tapped rectifier/transformer combination was used to provide this voltage, it could also provide the positive HT. As a further measure to protect the output valves, we could use a valve rectifier for the positive HT. If power was briefly interrupted to the amplifier, the valve rectifier would ensure that at the instant of power returning, the output valves would initially be biased off, but would gently be turned on as the rectifier warmed. Unfortunately, valve rectifiers drop more volts, so the positive HT is likely to be $\approx 300\,V$.

Although DC conditions do not require the HT for the first differential pair to be regulated, it is probably the best way of achieving a sufficiently low HT ripple. In order to ensure that the regulator does not drop out of regulation when mains voltage drops, we could set its output to $+270\,V$.

The first differential pair now has quite a low HT voltage, and the only way to maintain linearity and voltage swing is to reduce I_a. Reducing I_a allows V_a to swing closer to $0\,V$, which increases maximum output swing, and distortion is generally inversely proportional to maximum output swing. Slithering a transparent ruler over the anode characteristic curves resulted in an operating point of $V_a = 125\,V$ and $I_a = 2.9\,mA$ for each valve, using a $50\,k\Omega$ load resistor.

Valve matching

$2.9\,mA$ per triode is well below the ideal $8\,mA$, but the differential pair cancels the predominant 2nd harmonic distortion, and the voltage swing is very low, so this is tolerable, but the required distortion cancellation would benefit from matched valves in the differential pair.

Although the Loctal 7N7/14N7 valves tend to be quite well matched, individual triodes such as the 6J5GT allow far better matching. If we decided to standardize on single triodes, we would need 12 6J5GT for a stereo amplifier. The advantage of needing many valves of a given type is that the probability of finding matched pairs increases dramatically with the number of valves, so buying the required dozen valves offers a far better chance of finding matched pairs than buying two. It was the decision to use a dozen

6J5GT of complementary appearance to the 13E1 output valves that prompted the name 'Crystal Palace' for this amplifier.

In general, when two apparently identical valves have the same anode voltages in a differential pair, their gain is likely to be matched under those operating conditions. Thus, we can test 6J5GT in the amplifier by inserting a valve deemed to be the reference valve in one side of the differential pair, and sequentially test all the other valves against this reference. The valves with closest anode voltages are pairs.

The essential twiddly bits

We have made the broad brushstrokes, and chosen our stage topology, valve type, anode currents and load resistances. It is now time to get down to the nitty-gritty and ensure that those conditions are met. In order, we need to:

- Set the DC conditions of each stage by designing their constant current sinks.
- Consider thermal stability of the constant current sinks.
- Consider RF stability by including grid-stopper resistors and bypassing the HT supplies correctly.
- Design the HT regulators.

The cascode constant current sink and stabilization against mains variation

We know that $V_{gk} \approx -10\,V$ for the second differential pair, so we should be able to design a cascode that operates at this voltage, thus avoiding a subsidiary supply.

For reasons that will become apparent in a moment, we need a low reference voltage, so we will use an infrared LED, which we will bias from the 0 V rail via a large resistor. Because the resistor has $\approx 350\,V$ across it, it can only pass quite a small current to keep within its power rating. If we choose a 150 kΩ, 3 W resistor, it will pass 2.3 mA whilst dissipating 0.83 W. 2.3 mA would not normally be considered to be an ideal reference current for an IR LED because $r_{internal}$ rises significantly at lower currents (16.4 Ω at 2.3 mA, as opposed to 5.4 Ω at 10 mA). However, because the slope resistance helps compensate for mains voltage variations, this isn't a problem.

In order to consider the effect of mains variation, we will assume a 1% rise in mains voltage.

In the output stage, we want to hold I_a constant despite changing V_a, and find the V_{gk} that would oppose this change. This is effectively the definition of μ.

Since the output stage HT = 400 V, a 1% rise implies +4 V. For the 13E1, $\mu \approx 3.9$, so V_{gk} must fall by ≈ 1 V to combat the anode change.

A 1 V fall at an anode of the second differential pair would be caused by an increase in individual anode current of:

$$I = \frac{V}{R} = \frac{1}{50\,\text{k}\Omega} = 20\,\mu\text{A}$$

But there are two valves, so the tail current must increase by twice this, = 40 μA.

For our infrared LED $V_{ref.} = 1.10$ V at 2.33 mA, so the current programming resistor in the emitter circuit of the cascode must be:

$$R = \frac{V}{I} = \frac{1.10 - 0.7}{10\,\text{mA}} = 40\,\Omega$$

A change of 40 μA in the 40 Ω programming resistor would be caused by a change in voltage of $V = IR = 40\,\mu\text{A} \times 40\,\Omega = 1.6$ mV.

Assuming constant base/emitter voltage, $V_{ref.}$ must rise by 1.6 mV to combat the change in output stage current.

The 150 kΩ resistor passes 1% more current due to the 1% rise in mains voltage. It normally passes 2.33*r* mA, so the increase in current is 23.33*r* μA.

We now know the change in current and change in voltage across an unknown resistance, so we can find its value:

$$R = \frac{V}{I} = \frac{1.6\,\text{mV}}{23.33\,\mu\text{A}} \approx 68.6\,\Omega$$

The IR LED passes 2.33 mA, and contributes $r_{slope} = 16.4\,\Omega$, so we need $68.6\,\Omega - 16.4\,\Omega = 52\,\Omega$.

But 52 Ω drops 121 mV at 2.33 mA, so $V_{ref.}$ rises to 1.10 V + 0.121 V = 1.22 V.

Therefore the voltage across the programming resistor becomes 521 mV, and since it must pass 10 mA, its required resistance changes from 40 Ω to 52 Ω. See Fig. 6.43.

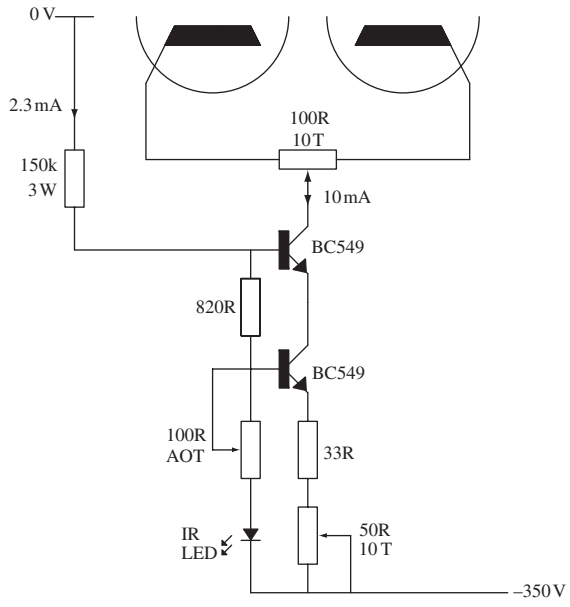


Fig. 6.43 *Biasing the second differential pair*

As can be seen, the values of the programming and compensating resistors interact, so we will need to use variable resistors and adjust them on test as follows:

1 Set the compensating resistor to its maximum value, set the programming resistor to its minimum value.
2 Adjust the programming resistor for correct output stage current.
3 Using a variac, raise mains voltage by 5%.
4 Adjust the compensating resistor to restore correct output stage current.
5 Restore correct mains voltage.

Repeat steps (2) to (5) until the output stage current variation with mains voltage is minimized.

The transistors do not need to withstand large voltages, so BC549 is ideal.

We know that we need to be able to adjust the relative voltages on the grids of the output valves to equalize individual anode currents, so a resistor between the cathodes of the second differential pair allows this variation. If

we use a 100 Ω resistor and notionally move the wiper to one extreme end, the resistor will pass the current of only one 6J5GT, which is ≈5 mA, so it will drop ≈0.5 V. The gain of the differential pair is ≈18, so there will be a grid to grid change of ≈9 V at the output valves. Because the wiper could be moved to the opposite end of the resistor, we could achieve the same voltage change but in the opposite direction. Thus, each valve can effectively have its V_{gk} varied ±9 V, which ought to be quite sufficient to achieve anode current balance.

The collector of the lower transistor hardly has to change voltage, so $V_{ce} = 2\,V$ is perfectly satisfactory for this transistor at 10 mA. Because the collector of the lower transistor is connected to the emitter of the upper transistor, V_{ce} is equal to the voltage between the two emitters. Because we drop 0.7 V across the base/emitter junction of both transistors, V_{ce} for the lower transistor is equal to the voltage between the two bases. When we put a resistor between the bases, we know that it passes the 2.33 mA sourced via the 150 kΩ resistor, so its required value is 2V/2.33 mA ≈ 820 Ω.

The 334Z constant current sink and thermal stability

According to the data sheet,[16] the current programming resistor for the 334Z can be calculated using:

$$R = \frac{227\mu V \times T}{I_{set}}$$

where T is the absolute temperature.

If we assume an ambient temperature of 300 K (27 °C), this simplifies to:

$$R = \frac{68.1}{I_{set(mA)}}$$

Thus, to set 5.8 mA, we need ≈12 Ω resistor. However, on test, 24Ω was required. Nevertheless, it was worth seeing the first equation because it reminds us that all electronics drifts with temperature. The usual reason for drift is the temperature dependence of V_{be} for silicon transistors, but this can usually be compensated by adding a silicon diode in the reference chain. The fundamental assumption is that the diode is at the same temperature as the junction producing the error, so the compensating diode should be glued to the offending device with epoxy adhesive, and the entire mass insulated from convection currents by a small expanded polystyrene shroud.

Sure enough, the data sheet gives a circuit that compensates for thermal drift, and simply requires that the additional resistor is ten times the programming resistor. See Fig. 6.44.

Having compensated the (uncritical) 334Z, we should consider compensating the cascode constant current sink because this *is* critical. The traditional method of compensating the cascode adds a silicon diode in series with the reference diode to compensate for the changing V_{be} of the lower transistor. The assumption is made that the reference diode has zero drift with temperature, and this is very nearly true if a 6.2 V Zener is used, but we have chosen to use an LED. Because the forward drop of the LED falls with increasing temperature, it already tends to compensate the transistor, so no extra components are required.

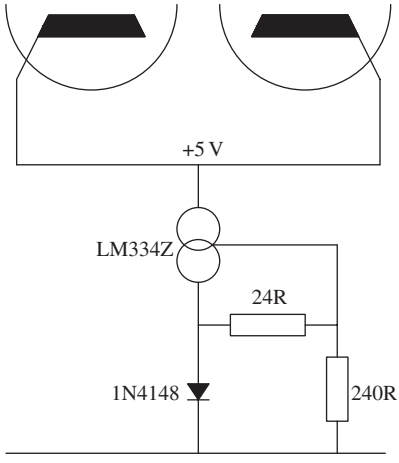
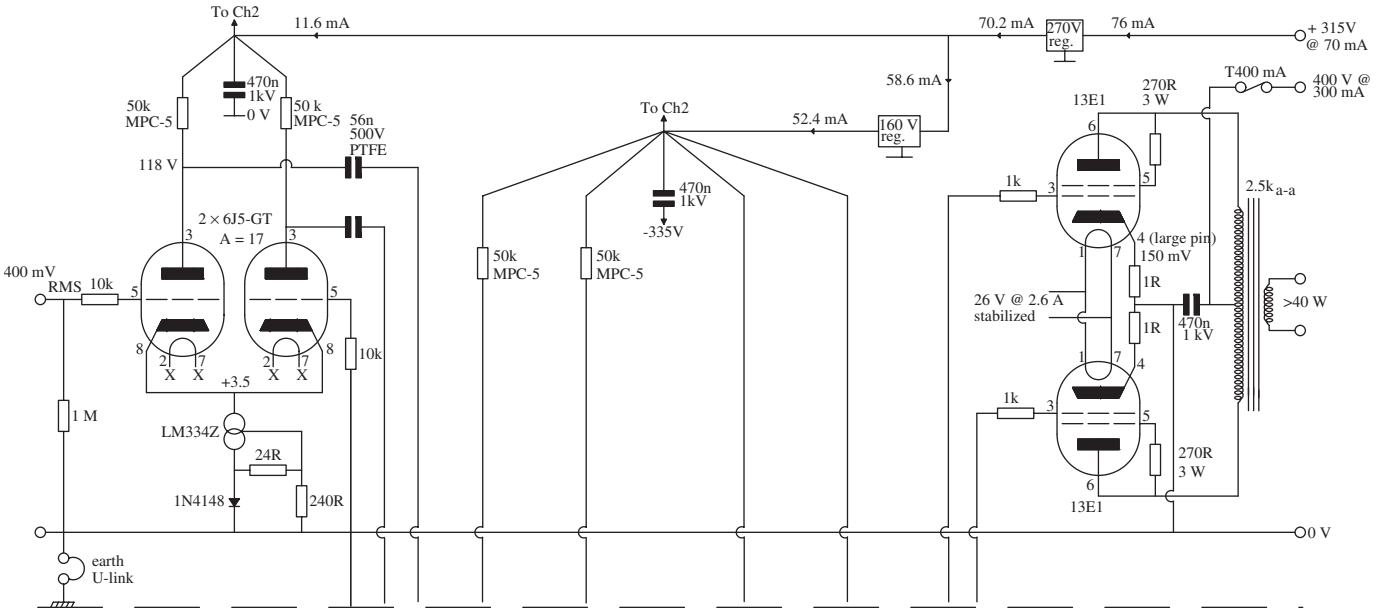


Fig. 6.44 *Compensating the 334Z against temperature variations*

HF stability

High *gm* valves such as the 13E1 are prone to parasitic oscillation, but this can be cured by grid-stopper resistors. Because cathode followers operate with 100% negative feedback, there is a danger of oscillation, so these need 10 kΩ grid stoppers. Differential pairs rarely need grid stoppers, but shorting a grid directly to ground is inviting RF oscillation (no damping whatsoever), so 10 kΩ grid stoppers could be fitted to the first differential pair.

Adding 10 kΩ grid stoppers to the second differential pair would reduce the $f_{-3\,dB}$ point due to Miller capacitance and source resistance from an


To Ch2
11.6 mA
70.2 mA
270V reg.
76 mA
+ 315V @ 70 mA
58.6 mA
T400 mA
400 V @ 300 mA
270R 3 W
13E1
50k MPC-5
470n 1kV
0 V
50 k MPC-5
56n 500V PTFE
118 V
To Ch2
52.4 mA
160 V reg.
470n 1kV
-335V
2 × 6J5-GT
A = 17
50k MPC-5
50k MPC-5
1k
2.5k a-a
4 (large pin)
150 mV
1R
26 V @ 2.6 A
stabilized
1R
470n 1 kV
>40 W
400 mV
RMS
10k
10k
1 M
LM334Z
24R
1N4148
240R
1k
270R 3 W
13E1
+3.5
earth
U-link
0 V

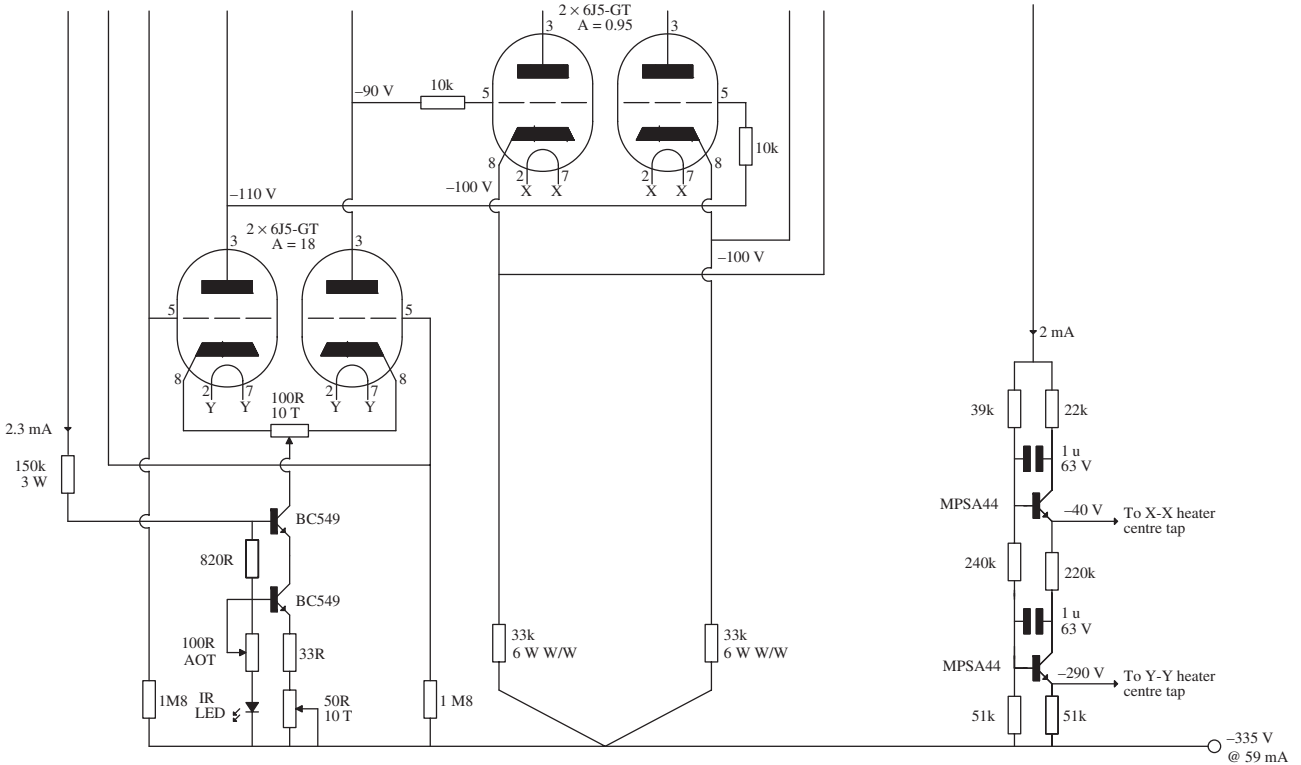


Fig. 6.45 *Final audio circuit of 'Crystal Palace' power amplifier with (measured) voltages*

acceptable 130 kHz to an unacceptable 60 kHz, so grid stoppers were omitted.

Another possible cause of HF oscillation is non-zero power supply impedance. To counter this problem, the output of the +160 V regulator should be a star point, and the −350 V HT should also feed a star point. A 470 nF capacitor can then be connected between these star points to ensure stability of the second differential pair and its associated cathode followers. Similarly, a 470 nF capacitor should be connected from the centre tap of the output transformer to the junction of the 1Ω current sense resistors in the output stage, and another 470 nF capacitor from the star point of the 270 V regulator to the bottom of the 1N4148 diode in the *334Z* constant current sink. Finally, we need a THINGY to set V_{hk} appropriately. THINGY design has been covered previously, so we now have our final audio design. See Fig. 6.45.

HT regulators

The second differential pair requires a 160 V regulator free from DC drift, so a Maida regulator is ideal. The first differential pair is not critical, but there is no real reason to use an alternative, so another Maida regulator is used. HT regulator design has been covered previously, and there is nothing exciting in these examples. Supplying the 160 V regulator via the 270 V regulator ensures that the 270 V regulator passes sufficient current to operate correctly. With two component changes, the same design is used for both the 270 V and 160 V supplies. See Fig. 6.46.

Stereo vs mass

The author began the metalwork and layout of the 'Crystal Palace' amplifier as a stereo amplifier *before* buying accurate scales. At 90 lb, this is not an amplifier for the faint-hearted.

The reason for building the amplifier as a stereo amplifier is that the totally balanced audio topology renders the amplifier insensitive to power supply noise. There is therefore no need to have separate left and right power supplies, and a considerable reduction in support circuitry can be achieved. Last but not least, the author had many of the parts to achieve a stereo amplifier, but a pair of mono chassis would have doubled the metalwork and required the purchase of a pair of large HT chokes. You might have a different opinion about the benefits of a stereo chassis.

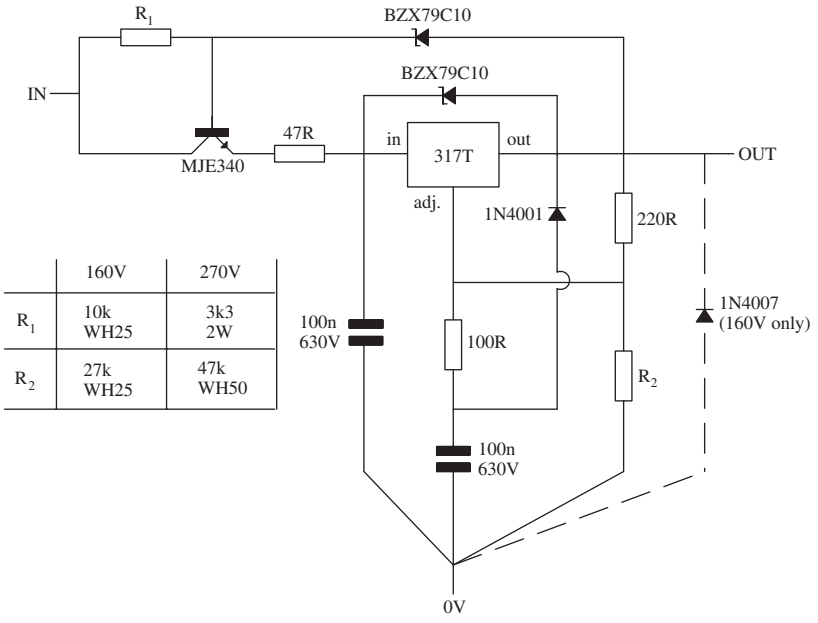


Fig. 6.46 *Regulator design*

Power supply design

Having decided on a stereo chassis, we need an HT supply capable of supplying 1 A at ≈400 V. A quick check with a spreadsheet revealed that a choke input HT supply would need a 2 H, 1.5 A choke and a 455 V_{RMS} mains transformer. The author took one look at his 1 H 1 A choke, and decided that an even larger choke was not acceptable. A capacitor input HT supply was therefore necessary. 1200 V fast recovery diodes are used for the bridge rectifier. In order to protect the 13E1 output valves, we can fit a thermal delay relay in the mains path to the output stage HT transformer.

The 13E1 heaters require 26 V at 2.6 A per channel, so a 2 × 25 V_{RMS} 300 VA toroid was chosen because a 160 VA transformer would have been marginal, and the 300 VA transformer cost no more than a 250 VA, yet was still sufficiently slim to fit inside the 2″ chassis. The regulator arrangements are standard, but the reservoir capacitors are deliberately small to reduce dissipation in the regulator.

Traditional centre-tapped HT transformers are intended for use with valve rectifiers, so only one half of the winding is in use at any instant. If we derive a positive *and* a negative supply, both windings are in use simultaneously, so we

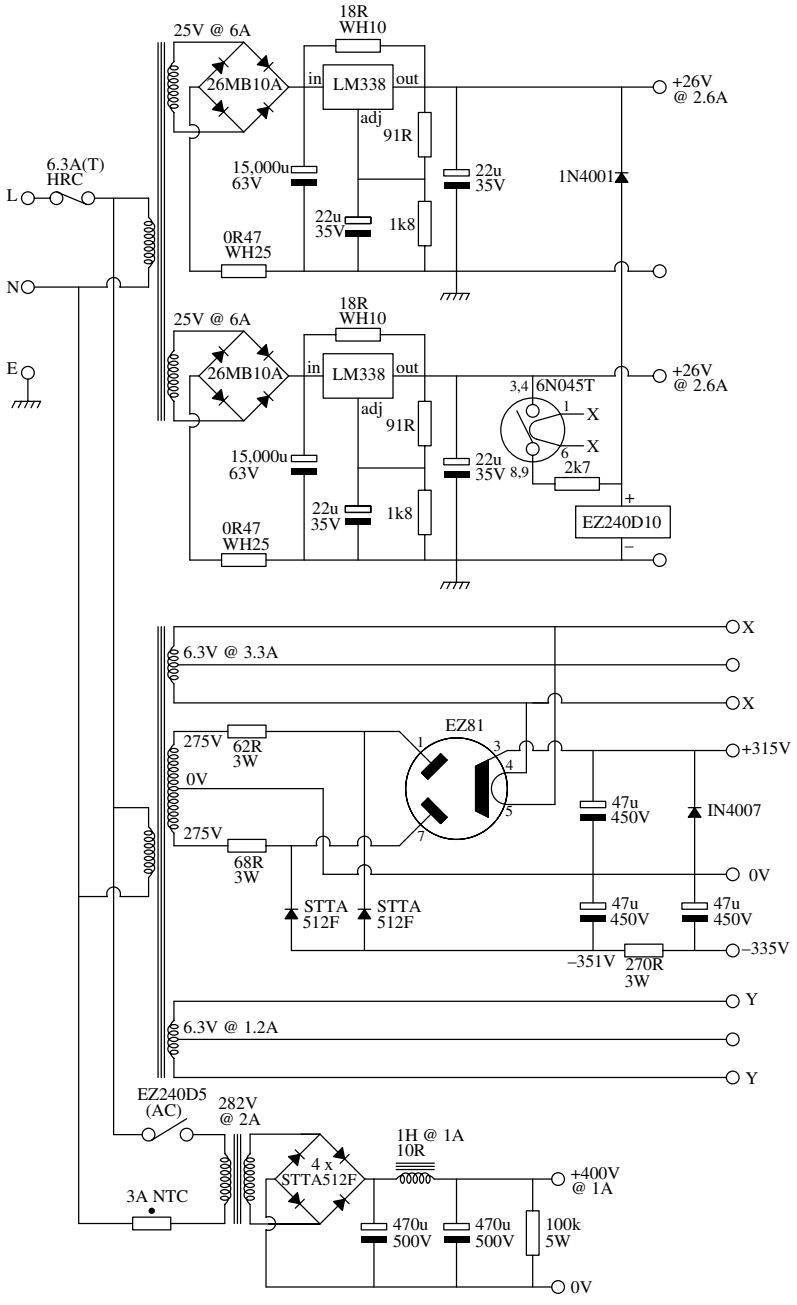


Fig. 6.47 *Main PSU design*

must be careful not to exceed the VA rating. The simplest way to ensure this is to say that the sum of the positive and negative currents must be less than the winding rating. Thus, if we need 78 mA for the positive supply, and 61 mA for the negative, the total current is 139 mA, and a 150 mA 275 V–0–275 V winding is fine. Fortuitously, the salvaged HT transformer also had the bonus of a pair of 6.3 V 4 A centre-tapped windings suitable for the driver valves, EZ80 rectifier and delay relay. See Fig. 6.47.

Global negative feedback and biassing

As shown, the amplifier does not have global negative feedback. If required, feedback could be taken from the amplifier output terminals to the grid of the first differential pair that is presently grounded.

One factor that you might like to consider is that feedback reduces the blocking margin. We saw earlier that the output stage could be overloaded by 10 dB before the second differential pair would suffer blocking. Clipping breaks the feedback loop, so the driver operates with its open loop gain, making blocking much more likely.

As an example, suppose that we applied 6 dB of global negative feedback to the amplifier. Next, we apply a sine wave to the amplifier and increase its level in 1 dB steps until the amplifier produces its maximum undistorted output. We then increase the input by a further 1 dB. The output stage clips, the feedback loop is broken, and the gain of the amplifier promptly increases by 6 dB. We have now used up 7 dB of our blocking margin.

On the author's chassis, the 13EI's could not safely dissipate 95W each without adding four noisy fans. Thesefore, bias current was backed off to 150 mA per valve, resulting in class AB.

'Daughter of Beast'

Neat, sweet and petite, this updated direct drive amplifier for electrostatic headphones is derived from the 'Beast' power amplifier's driver circuit (described in the 2nd edition), hence its name. The amplifier was designed to produce a maximum undistorted output of 200 V_{RMS} at each terminal, plus 550 V bias from 10 MΩ, in order to match the principal electrical characteristics of the Sennheiser HEV70 it was designed to replace. See Fig. 6.48.

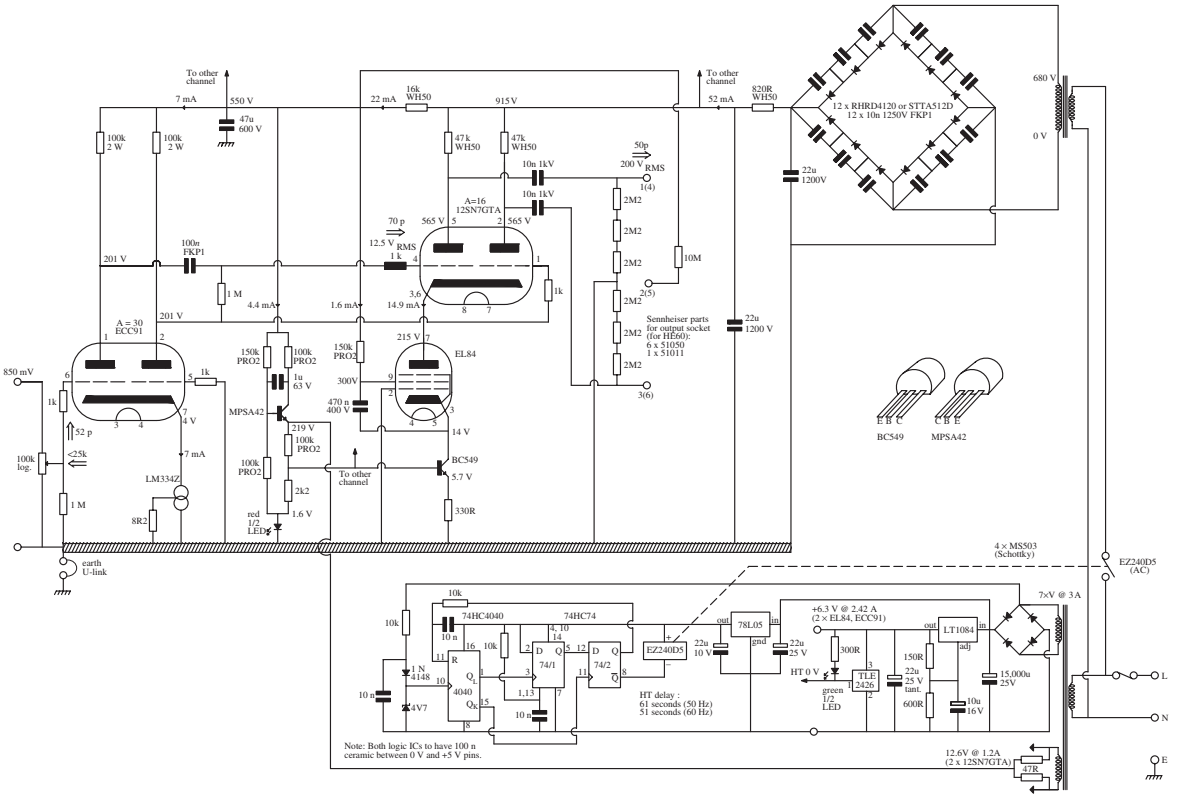


Fig. 6.48 *'Daughter of Beast' – a direct-drive electrostatic headphone amplifier*

The amplifier is ultimately intended to be driven directly from a balanced $2\,V_{RMS}$ DAC (modern 1-bit DACs incorporate a digital phase splitter producing a differential output which is normally converted to single ended by external circuitry, which seems a waste).

The original version of this design used a 6BX7 as the output valve, but experience showed that these rather variable valves need to be selected to find a pair of low distortion triodes in the same envelope, so the design has been modified to use the ($V_{a(max.)} = 450\,V$) 12SN7GTA, but the true common cathode ECC91 has been retained. In any valve, the cathode deteriorates throughout its life, but as the same physical cathode is in use for both sections in the ECC91, we can reasonably expect initial balance to be maintained – a worthwhile advantage. Sadly, other electrically 'common cathode' valves, such as the E90CC and E92CC, actually have individual cathodes connected internally.

The operating point of the 12SN7GTA is critical to maximize output voltage, so a transistor in the cathode makes a hybrid BJT/pentode cascode, which allows I_k to be determined purely by semiconductor parameters, enabling valve replacement without having to reset I_k.

In a desperate effort to reduce the amplifier's size and weight, a conventional capacitor input HT supply has been used with minimal smoothing, so this should be checked to ensure that the simplification is permissible.

Calculating output hum due to the HT supply

The principles of the following argument can be applied to other areas, so the topic is covered in some detail. Additionally, an interesting result emerges, which is well worthy of further thought.

The ripple voltage across the $22\,\mu F$ reservoir capacitor is:

$$V = \frac{It}{C} = \frac{0.053 \times 0.01}{22 \times 10^{-6}} \approx 2\,V_{pk-pk}$$

This is then attenuated by the following $820\,\Omega$ and $22\,\mu F$ network. If we consider 100 Hz (double mains frequency) to be the only significant component of the ripple sawtooth waveform, then to a very close approximation, the loss is:

$$\text{Loss} \approx \frac{X_{C(100\,Hz)}}{R} \approx \frac{72}{820} \approx 0.088$$

The ripple voltage fed to the output stage is therefore $\approx 175\,\mathrm{mV_{pk-pk}}$.

As mentioned in Chapter 2, the attenuation of HT noise at the anode is quite low because for a differential pair, r'_a looking into one terminal is quite high:

$$r'_a = R_L + 2r_a$$

Substituting this into the potential divider equation, the loss of the potential divider formed by r'_a and R_L is:

$$\text{Loss} = \frac{R_L + 2r_a}{R_L + R_L + 2r_a} = \frac{R_L + 2r_a}{2(R_L + r_a)}$$

Provided that $R_L >> r_a$, this tends to $\frac{1}{2}$, or $-6\,\mathrm{dB}$. We would therefore expect to see $\approx 90\,\mathrm{mV_{pk-pk}}$ of ripple on each anode of the differential pair.

At this point, we must be very careful to compare like with like, so when we compare the ripple to the signal voltage, the signal voltage must be expressed as V_{pk-pk}. We must also ensure that we are comparing voltages between the same terminals. Because the output is taken from between the two anodes (which are inverted with respect to one another), we see double the voltage seen to ground from one anode. The maximum peak-to-peak voltage between one stator plate and the other of the headphones is therefore:

$$V_{pk-pk(max)} = 200 \times \sqrt{2} \times 2 \times 2 \approx 1100\,\mathrm{V_{pk-pk}}$$

The power supply ripple voltage may be related to maximum output voltage:

$$\frac{\text{ripple voltage}}{\text{maximum output voltage}} = \frac{0.09}{1100} \approx -82\,\mathrm{dB}$$

Ideally, the headphones only respond to the *difference* in voltage between one anode and the other, and would therefore completely reject any common-mode ripple. The practical headphones' rejection of common-mode signals is determined by the equality of the insulating spacer thickness separating the diaphragm from either stator. Typically, the spacers might be made of polystyrene sheet, which can vary in thickness across the sheet, so two spacers stamped from a sheet could be of differing thicknesses, perhaps as much as 3% in error, expressed in dB:

$$\mathrm{dB} = 20\log(0.03) \approx -30\,\mathrm{dB}$$

It is therefore reasonable to suppose that the headphones are only likely to reject common-mode ripple by 30 dB, although they could very well be much better. We can now see that the audible power supply hum is likely to be ≤ -112 dB below maximum output level, which is probably satisfactory.

However, we should not be complacent about low levels of hum, because hum is at a fixed frequency, and the eye, or ear, brain combination is $\approx 15-20$ dB more sensitive to fixed pattern noise compared to random noise. A practical example of this difference in sensitivity is that when analogue music circuits were commonly provided by British Telecom, they would allow the customer to reject a circuit with random noise worse than -43 dBu, but single tone interference could be rejected at -60 dBu (17 dB lower in level).

Using the previous argument, the power supply hum could be considered to be equivalent to random noise ≈ 17 dB louder, which would be 95 dB below maximum level.

A digital meander

A correctly dithered 'n'-bit digital system (perhaps CD) through A–D and D–A has a theoretical (unweighted) Signal to Noise (S/N) ratio referred to maximum output of[17]:

$$\mathrm{S/N} = 6n(\mathrm{dB}) - 3\,\mathrm{dB} = 6 \times 16 - 3 = 93\,\mathrm{dB}$$

The 3 dB factor is the unavoidable dither (noise) that should be added to the A–D's quantizer to linearize quantization and avoid noise modulation of music signals, although carefully chosen digitally synthesized signals may not require full dithering, allowing an improvement in S/N ratio of 4.76 dB to be achieved.

To summarize the previous digital arguments, the maximum unweighted S/N ratio of a digital system is defined by the channel, rather than the converters, so a 24-bit D–A on a 16-bit channel (such as CD) simply allows a closer approximation to the theoretical S/N ratio. However, it is possible to gain a purely subjective improvement by manipulating the frequency distribution of the quantizing noise using a technique known as noise weighting. Currently, this allows a usable improvement of ≈ 18 dB, measured via an A-weighting filter, which should also result in a subjective improvement of ≈ 18 dB, so a noise weighted 16-bit channel could have an S/N of 111 dB(A).

Practical significance of the digital meander

The full output of a balanced output CD player easily drives the amplifier to its maximum output level, so the random noise from a standard CD should just mask the power supply hum, but a true 20-bit signal (perhaps DVD-A) or noise weighted signal, applied through a superior converter would be compromised. This design would therefore ideally like to see headphones with rather better mechanical construction than postulated in order to ensure acceptable hum levels – unusually, a normally electronically determined characteristic is now critically dependent on the mechanical construction of a transducer.

Heat

Changing the output valves from 6BX7 to 12SN7GTA reduces output valve heat dissipation by almost 40%, greatly assisting cooling. Nevertheless, some rules were stretched in the mechanical construction of this amplifier.

- The chassis is rather small. However, a larger chassis would not have been tolerated – this is a headphone amplifier, after all!
- The output valves are rather close together. Fortunately, they are not operated at maximum power rating for the envelope, but cooling is still an issue.

Fig. 6.49 *The author's prototype*

A small chassis is permissible *if* the cooling problems can be solved. The 47 kΩ WH50 anode load resistors and LT Schottky diodes are thermally bonded to the chassis in order to minimize air temperature within the chassis. The valves are mounted on perforated aluminium sheet, whilst the bottom cover of the amplifier is entirely perforated, and the amplifier is supported by loudspeaker carpet piercing spikes, allowing a free flow of cooling air past the valves and the components immediately beneath. See Fig. 6.49.

References

1 Hafler, David and Keroes, Herbert I. (1951) An Ultra-Linear Amplifier, *Audio Engineering*, November pp. 15–17.

2 British Patent No. 496,883. (1937) Improvements in or Relating to Thermionic Valve Amplifying Circuits. Alan Blumlein.

3 Futterman, Julius (1956). A Practical Commercial Output Transformerless Amplifier. *Journal of the Audio Engineering Society*, October. Also US Patent No. 2,773,136 (Dec. 1959).

4 Schmitt, O. H. (1938) Cathode Phase Inversion, *J. Sci. Inst.*, 15, 136, pp. 100, 101.

5 US Patent No. 1,913,449. Kobayashi, M. (1929).

6 The Williamson Amplifier. A collection of articles reprinted from *Wireless World*, April 1947, May 1947, August, October and November 1949, December 1949, January 1950, May 1952.

7 Hafler, David and Keroes, Herbert I. (1952). Ultra-Linear Operation of the Williamson Amplifier, *Audio Engineering*, June.

8 Mullard (1959). Tube Circuits for Audio Amplifiers. Now reprinted by Audio Amateur Press, Peterborough, New Hampshire, USA (1993).

9 Walker, Peter (1993) Private communication.

10 Williamson, D. T. N. and Walker, P. J. (1954) Amplifiers and Superlatives, *Journal of the Audio Engineering Society*, April, Vol. 2, No. 2.

11 General Electric Company Ltd. (1957). *An Approach to Audio Frequency Amplifier Design*. Now reprinted by Audio Amateur Press, Peterborough, New Hampshire, USA (1994).

12 Colloms, Martin (1997) *High Performance Loudspeakers* 5th ed., pp. 285–302. Wiley.

13 Hedge, L. B. (1956) Cascade AF Amplifier, *Wireless World*, June.

14 Langford-Smith, F. (1957) *Radio Designer's Handbook*, 4th ed. Iliffe.

15 Philips QE 5/40 valve data sheet. 12 December 1958.

16 LM134–LM234–LM334 data sheet. SGS-Thomson. October 1997.
17 Watkinson, John (2000) *The Art of Digital Video*, 3rd ed. Focal.

Further reading

Borwick, John (ed.) (1994) *Loudspeaker and Headphone Handbook*, 2nd ed. Focal.
Linsley Hood, John (1997) *Valve and Transistor Audio Amplifiers*. Newnes.
Manley, David. (1994) *The Vacuum Tube Logic Book*, 2nd ed. Vacuum Tube Logic.

7

The pre-amplifier

The pre-amplifier receives a variety of signals at different levels, processes them, and passes them to the power amplifier. It should do this without adding noticeable noise or distortion and should be convenient to use.

The block diagram of a typical pre-amplifier is shown in Fig. 7.1.

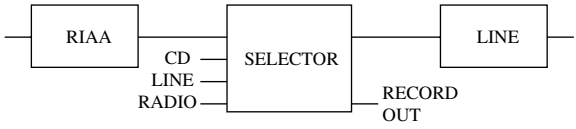


Fig. 7.1 *Block diagram of pre-amplifier*

Working back from the output, we have a line stage which provides a limited amount of gain; it may be designed to drive long cables, and it may even include tone controls. This is preceded by the volume control and input selector. One of the sources to the selector switch is an RIAA equalized disc stage, although many modern pre-amplifiers neglect this stage because it is so difficult to design. The excuse for this is that the LP is obsolete. This may be so, but there are many irreplaceable recordings on LPs that need to be played, so a comprehensive pre-amplifier should include a disc stage.

The line stage

Determination of requirements

The pre-amplifier has to provide only a very limited output voltage to the power amplifier, as even the most insensitive power amplifiers do not require

more than 4 or 5 V_{RMS} to drive them into clipping. This means that we are concerned only with linearity and not headroom in this stage. The stage must drive the capacitance of the cable without excessive loss at high frequencies, and it may need to be able to drive power amplifiers of lower input impedance, such as transistor amplifiers. Since the stage is preceded by the volume control, the power amplifier amplifies *all* of the stage's self-generated input noise, so we must ensure that this noise is minimized.

These requirements dictate a low gain, coupled with low output impedance which ought to be quantified.

Low capacitance screened cable typically has a capacitance of ≈100 pF per metre. To avoid inducing hum from the mains transformers into the preamplifier, we probably need to separate them by one metre. By the time we allow for the routing of the cable between the line stage and the input plugs of the power amplifier, we will probably use about 1.5 m of cable, which is equivalent to 150 pF. The power amplifier has input capacitance, so we ought to allow ≈20 pF for a valve amplifier, and ≈200 pF for a transistor amplifier. This means that as an absolute minimum, we should be able to drive 170 pF, but 1 nF would allow for longer cables and transistor amplifiers with higher input capacitance.

The source impedance and cable shunt capacitance form a low-pass filter whose $f_{-3\,dB}$ cut-off frequency we can calculate from:

$$f_{-3\,dB} = \frac{1}{2\pi CR}$$

However, we would like the high frequency roll-off within the audio band to be far less than 3 dB, so we need to know what $f_{-3\,dB}$ corresponds to a given amount of loss at a given frequency, which we can find from the following formula:

$$f_{-3\,dB} = \frac{f_{(dB\ limit)}}{\sqrt{\frac{1}{10^{\frac{dB}{10}}} - 1}}$$

where: $f_{(dB\ limit)}$ = the outermost frequency of interest
dB = the deviation from flat response at the above frequency

As an example of using this equation, we find that for 0.1 dB roll-off at 20 kHz, we require $f_{-3\,dB}$ = 131 kHz. It should be noted that this formula is

only valid for a single HF CR or LR network. For a single LF CR or LR network, the formula becomes:

$$f_{-3\,\mathrm{dB}} = f_{(\mathrm{dB\ limit})}\sqrt{\frac{1}{10^{\frac{\mathrm{dB}}{10}}} - 1}$$

Using this formula, 0.1 dB roll-off at 20 Hz corresponds to $f_{-3\,\mathrm{dB}} = 3$ Hz.

For quick reference, the following table describes the response of 6 dB/octave high-pass and low-pass filters in terms of multiples and sub-multiples of their cut-off frequency:

HPF	$f/7$	$f/6$	$f/5$	$f/4$	$f/3$	$f/2$	f	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$
dB	−17.0	−15.7	−14.1	−12.3	−10.0	−7.0	−3.0	−0.97	−0.46	−0.26	−0.17	−0.12	−0.09
LPF	$7f$	$6f$	$5f$	$5f$	$3f$	$2f$	f	$f/2$	$f/3$	$f/4$	$f/5$	$f/6$	$f/7$

As an example, if we investigate a high-pass filter one octave above its cut-off frequency ($2f$), we see that the response is −0.97 dB.

Once we cascade stages, the HF and LF cut-offs begin to move towards the mid-band, which is why it was suggested in Chapter 2 that 1 Hz was a good choice for an LF cut-off.

For 'n' stages, each with identical LF cut-off, the cut-off frequency of each *individual* stage is related to the cut-off of the composite amplifier by:

$$f_{-3\,\mathrm{dB(individual)}} = f_{-3\,\mathrm{dB(composite)}} \cdot \sqrt{2^{\frac{1}{n}} - 1}$$

Applying this formula to a three-stage ($n = 3$) capacitor coupled amplifier, we now find that our 3 Hz cut-off for the entire amplifier requires each stage to have a 1.5 Hz cut-off. The traditional value of 0.1μF coupling capacitor into 1 MΩ grid leak gives a cut-off of 1.6 Hz.

It is far more difficult to control HF cut-offs than LF cut-offs, so it is highly unlikely that we would ever have a series of amplifiers each with identical HF cut-offs, but if we did, the following formula could be used:

$$f_{-3\,\mathrm{dB(individual)}} = \frac{f_{-3\,\mathrm{dB(composite)}}}{\sqrt{2^{\frac{1}{n}} - 1}}$$

If we assume that our only significant HF cut-off is caused by cable capacitance loading the output resistance of the line stage, our earlier formula shows that for 0.1 dB loss at 20 kHz, driving our 170 pF minimum capacitance, we need an output resistance of $< 7\,k\Omega$, and preferably a good deal less.

We now need to make some assumptions about power amplifier input resistance and sensitivity.

Traditional power amplifiers had input resistance of 1 MΩ or more; this is a very good resistance, because it allows a low value of coupling capacitor from the pre-amplifier. 47 nF almost meets our 20 Hz 0.1 dB criterion, but 100 nF is better. Sensitivity is ≥ 1 V for the better amplifiers.

More sensitive amplifiers should be driven by a passive 'pre-amplifier' consisting simply of switches, volume control and a very short cable, or they should be redesigned to be less sensitive. In the 1960s, *power* amplifiers often had volume controls, selector switches and even tone controls on their front panels. This trend is now being repeated, as the LP, which was the prime reason for a separate pre-amplifier, is now a specialist preserve. The minority requirement for an RIAA stage is now generally met by stand-alone designs.

Modern valve power amplifiers tend to have a sensitivity of ≈ 500 mV, and an input resistance of 100 kΩ, requiring a coupling capacitor >470 nF. This low resistance seems to be because less care is taken over the heater wiring, which then induces a hum current into the input wiring of the first stage. The current develops a voltage across the grid leak resistor, which results in audible hum if the amplifier is not shunted by the low resistance source of a pre-amplifier. To avoid this, the grid leak is reduced to 100 kΩ, which gives a 20 dB improvement in hum when the input of the amplifier is left open circuit.

Transistor amplifiers also tend to have sensitivity of ≈ 500 mV, but have an even lower input resistance $\approx 10\,k\Omega$, which would require >4.7 μF coupling capacitor from the pre-amplifier.

A 4.7 μF 400 V capacitor is not only far more expensive than a 100 nF capacitor, but it is also an inferior component. Rather than trying to find an acceptable 4.7 μF component, we would do far better to attack the problem from the other end, and increase the input resistance of the power amplifier.

This is actually very easy for a valve amplifier. We simply replace the input grid leak resistor with a 1 MΩ resistor, and attend to the heater wiring if this generates a hum problem. This usually means twisting the heater wiring and pushing it into the corners of the chassis, but we may need to resort to powering the input valve heaters from a simple stabilized DC supply.

Many classic power amplifiers were designed to be too sensitive by modern standards. We could reduce sensitivity by fitting a potential divider near the

input valve, but a better solution is to reduce the gain of the input stage, and adjust global negative feedback to suit. The second solution is easily implemented, because the usual reason for excessive gain is that the input valve is an EF86 pentode. If we strap the valve as a triode (using the original anode load resistor) this reduces the gain of the stage by a factor of four, and if we keep the feedback factor constant, then the gain of the final amplifier will also be reduced by a factor of 4.

Theoretically, the cathode bias resistor should be recalculated, but in practice, the value need not be changed greatly. Because of the cathode feedback, the anode resistance of the valve used as a triode is high, and the output resistance of the stage is broadly similar to that using a pentode, therefore, compensation components in the anode circuit need not be changed. However, the global feedback network from the output of the amplifier to the input stage needs to be changed to maintain the correct level of feedback with the new gain. This could be done rigorously by calculating the old level of feedback, and recalculating for the new, but a rough method is to simply divide the resistor values by 4 (corresponding to the gain change from pentode to triode). Similarly, any capacitor value must be multiplied by 4 to maintain the correct RC time constant. See Fig. 7.2.

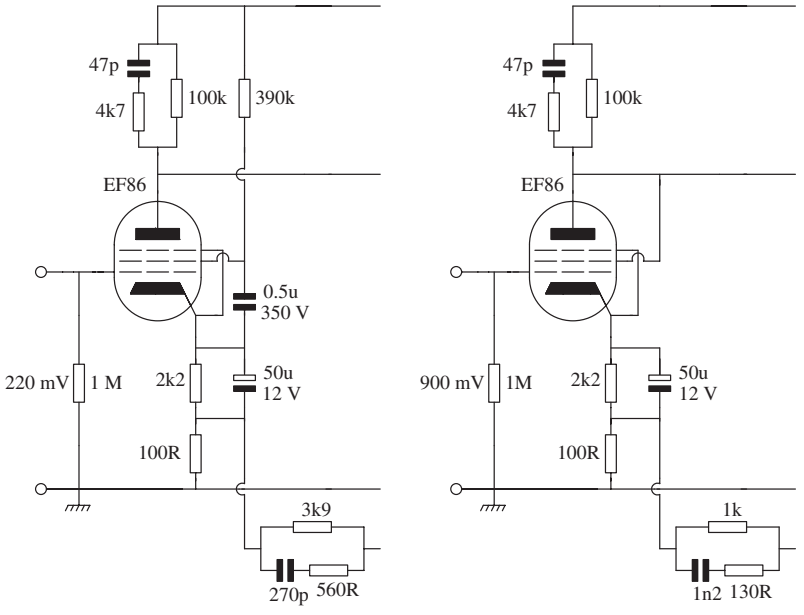


Fig. 7.2 *Modifying a typical pentode input stage for triode operation to reduce noise*

Attempting to raise the input resistance of a transistor amplifier without adding a buffer will probably result in a large DC offset at the output.

A traditional line stage

In general, valve pre-amplifiers should be designed to drive 2 V into 1 MΩ, even if it means modifying the power amplifier to achieve this match.

Older signal sources (particularly tuners) tend to produce only 250 mV of signal. If we must have a sensitivity of ≈250 mV at the input of the line stage, and the output is to be 2 V, we now know that we need $A_v = 8$. It is a good idea to have 3 dB to 6 dB more gain than this, to allow for unusually low recording levels, so $A_v = 12$ would be fine.

The stage will be preceded by a volume control, which we will investigate fully later, but for the moment, we can simply state that it will be a 100 kΩ logarithmic potentiometer, whose maximum output resistance is 25 kΩ.

This maximum output resistance may easily be verified by moving the wiper to the *electrical* mid-position of the track. The resistance to each end must be half the total resistance, and assuming zero source resistance, each end is at AC ground. Looking back into the potentiometer, we see the two halves in parallel, and therefore the output resistance is equal to the total resistance of the potentiometer divided by 4. If the wiper is at either end of the track, output resistance will be zero because it is connected either directly to ground or via the (zero resistance) source. Maximum output resistance therefore occurs when the wiper is as far away from *each* end as possible, which is the centre position.

The question of potentiometer output resistance is crucial, because it forms a low-pass filter in conjunction with the input capacitance of the line stage. Using our earlier argument of 0.1 dB HF loss at 20 kHz, when the volume control is at its maximum output resistance of 25 kΩ, we see that the maximum allowable input capacitance of the line stage is ≈50 pF.

If the input sensitivity of the stage is ≈170 mV (250 mV − 3.5 dB), and we want a signal to noise ratio of ≥100 dB, then the self-generated noise of the stage *referred to the input* will be 170 mV − 100 dB = 1.7 μV. Even the EF86 pentode can manage 2 μV on a good day, so we certainly ought to be able to better this using triodes.

Achieving the requirements

A good design is a *simple* design, so we should check to see if a common cathode triode amplifier could do the job. See Fig. 7.3.

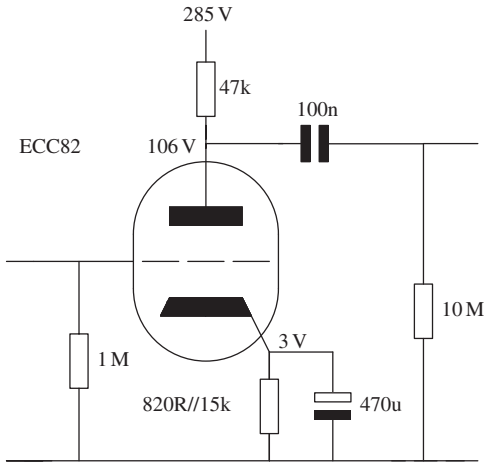


Fig. 7.3 *ECC82 common cathode amplifier as pre-amplifier output stage*

Once the stage topology has been decided, we can choose which valve is most suitable.

	Ideal	**ECC82**	***SN7**	***N7**	**ECC82 μ-follower**
A_v	12	15.5	17.5	17.5	19
r_{out}	$\not>7\,k\Omega$	7.7 kΩ	7.8 kΩ	7.8 kΩ	1 kΩ
C_{in}	$\not>50\,pF$	30 pF	77 pF	60 pF	35 pF

None of the proposed valves quite matches the ideal specification, but the ECC82 comes closest, and would be the obvious choice if it were not for the fact that the author's recent tests have shown that it produces ≈13 dB more irreducible distortion than a triode from the *SN7/*N7 family.

Although neither the *SN7 nor *N7 quite achieves the required r_{out}, the main problem is their C_{in}. Predicted C_{in} for the Loctal *N7 is only slightly higher than ideal, making $f_{-3\,dB} = 106$ kHz rather than 131 kHz, which would cause a loss of 0.15 dB at 20 kHz rather than 0.1 dB, but this loss might be deemed to be acceptable.

If we insist on using a simple common cathode stage, we have a choice between a linear valve with slight HF loss, or a valve with poorer linearity that almost meets all the specifications. Alternatively, we could incorporate the ECC82 into a μ-follower, which would significantly reduce its distortion and output impedance. See Fig. 7.4.

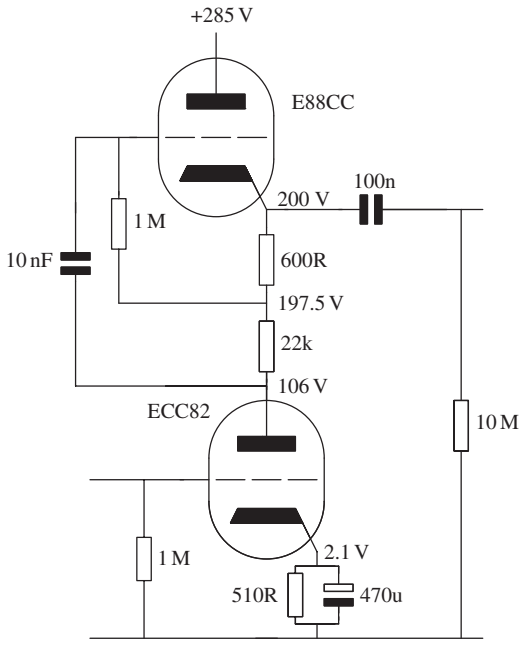


Fig. 7.4 *μ-follower as pre-amplifier output stage*

The μ-follower requires an elevated heater supply for its upper valve, but allows the required performance to be comfortably met. There is no 'correct' answer here. In 1993, the author used the ECC82 because it was cheap, simple and he had some. The Internet has made valves far more easily available, so faced with the same choice today, you might use the 7N7 with a low value of R_L (to reduce gain and C_{in}). Or, you might have an elevated heater supply ready and waiting, making the μ-follower a viable choice. There are many ways of skinning cats.

In the previous examples, we chose a valve and let it dictate the gain of the stage. In some instances, we may not be prepared to allow that to happen, and we will want to choose the gain.

Forcing the required gain

We could lower the gain by adding a potential divider at either the input or the output. Added at the output, it increases output resistance, which we cannot allow, whereas added at the input, it reduces signal level, but hardly affects noise, thereby degrading the S/N ratio.

Our only remaining option for adjusting gain is negative feedback. We can certainly reduce the gain of a single valve by removing the cathode bypass capacitor, causing series feedback, but this dramatically raises output resistance, and is therefore not allowed. In order to keep output resistance low, the feedback must be parallel derived, although it may be series or parallel applied. See Fig. 7.5.

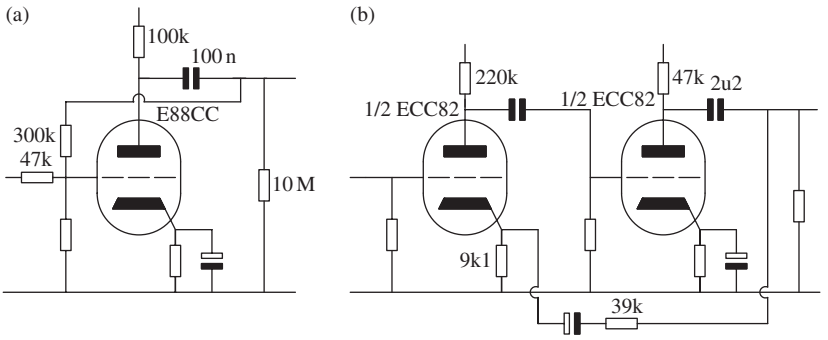


Fig. 7.5 *Shunt feedback v series feedback*

Both examples use parallel derived feedback to reduce output resistance, but in Fig. 7.5a, the feedback is parallel applied, whereas in Fig. 7.5b, it is series applied. Both solutions have problems, most of which are linked to the allowable value of the feedback resistor. The feedback resistor is effectively in parallel with the anode load, and is fed by the output coupling capacitor. A low value of feedback resistor requires a larger coupling capacitor and reduces the (open loop) gain of the stage by reducing the AC anode load resistance. A reasonable first choice for the feedback resistor might be $3 \times R_L$.

For the single valve invertor with $R_L = 100\,k\Omega$, we might want a final gain of 5, which would require a series input resistor of $47\,k\Omega$, and a feedback resistor of $300\,k\Omega$. The input resistance of this stage is $\approx 47\,k\Omega$, so it would heavily load a $100\,k\Omega$ volume control. We could increase both resistors, making the series input resistor $270\,k\Omega$, and the feedback resistor $1.6\,M\Omega$, thus reducing the loading on the volume control and further reducing the loading on the output stage. Unfortunately, this stratagem compromises noise performance because the series resistor generates noise, and the input noise of the grid leak resistor is no longer shunted by the source resistance of the volume control.

Having dismissed the invertor, we turn to the two valve stage, which has feedback taken to the cathode of the input stage. The grid does not have

a series resistor, and the cathode resistance will be quite low, so noise due to the input circuitry will not be compromised, although noise due to R_L will no longer be shunted. (See later in this chapter.)

Having individual input and output valves gives us the opportunity to optimize each stage. V_1 has been biased for low anode current in order to maximize the value of R_k (we will see why this is necessary in a moment), whilst V_2 has been biased for low r_a and good linearity. The capacitor in series with R_{fb} prevents DC from V_1 cathode from appearing on the output.

Total gain is ≈ 200 before feedback. If, as before, we need a total gain of 5, then we can find β the feedback fraction, from:

$$\begin{aligned}\beta &= \frac{1}{A} - \frac{1}{A_0} \\ &= 0.195\end{aligned}$$

The following approximation ignores the feedback current in R_k due to V_1, and also ignores the loading of r_k on the feedback network, but it is sufficiently accurate for our present purpose.

$$\begin{aligned}R_{fb} &\approx R_k\left(\frac{1}{\beta} - 1\right) \\ &\approx 39\,\text{k}\Omega\end{aligned}$$

This is a significant loading of the output stage, which decreases the open loop gain of the stage and increases distortion *before* the feedback is applied.

The loading of R_{fb} is always a problem in amplifiers of this type, and is reduced by increasing the required gain (reduces β), or by increasing R_k. Alternatively, we can accept the low value of R_{fb}, and arrange to be able to drive it by adding a cathode follower to the output of the amplifier. See Fig. 7.6.

We have now achieved our objective, but we needed three valves and an extra heater supply for the cathode follower, and we have not even considered the problem of ensuring stability of the completed amplifier. There is much to be said for using a gain structure whose gain is dictated by valve choice.

The volume control

The volume control is not simply a 100 kΩ potentiometer found at the bottom of the scrapbox. It is an essential part of a pre-amplifier and should be treated with the same care as any other part of the design.

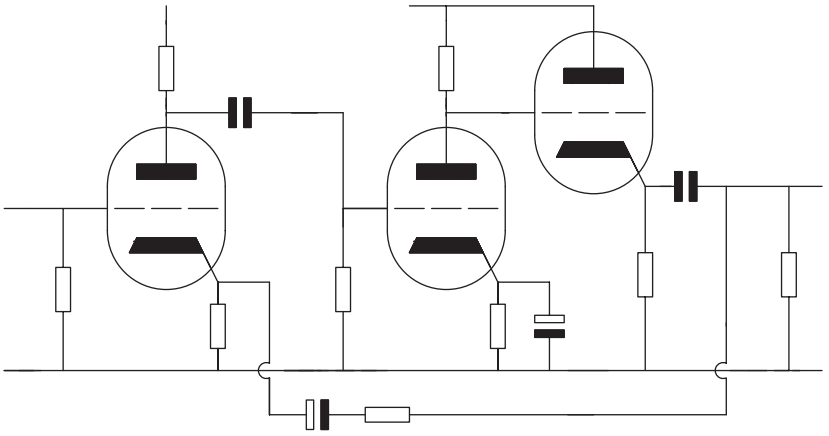


Fig. 7.6 *Adding a cathode follower to allow series feedback*

The human ear has a logarithmic response to sound pressure level, so if we want a volume control that has a uniform perceived response to adjustment throughout its range, we need a logarithmic potentiometer. This is the root cause of all our problems.

It is not a problem to make a linear potentiometer. We simply deposit a strip of carbon of uniform width and thickness onto an insulator, put terminals at each end, and arrange for a contact to scrape its way along. If we don't bother with a casing, it is known as a skeletal type. In an attempt to produce a logarithmic law, the coating thickness is made variable, then, in deference to audio sensibilities, a pressed metal screening can is fitted, and two potentiometers are ganged together on one shaft onto which we can fit a big, shiny, spun aluminium knob. Making the coating thickness continuously variable would be expensive, so the logarithmic law is approximated by a series of straight lines. See Fig. 7.7.

It is amazing how good a fit to the ideal logarithmic curve can be made using only four different resistance tracks, but it will come as no surprise to learn that this still results in steps in the response as the knob is rotated. We also expect the mechanically linked potentiometers to produce identical levels of attenuation all the way from 0 dB to 60 dB. Some of them are remarkably good, but the carbon track potentiometer's natural habitat is buried in the undergrowth of an old television.

Law faking

One useful fudge is to use a conductive plastic linear potentiometer with a law faking resistor added between wiper and earth. This does not produce a true logarithmic law, but it is far better than a pure linear potentiometer.

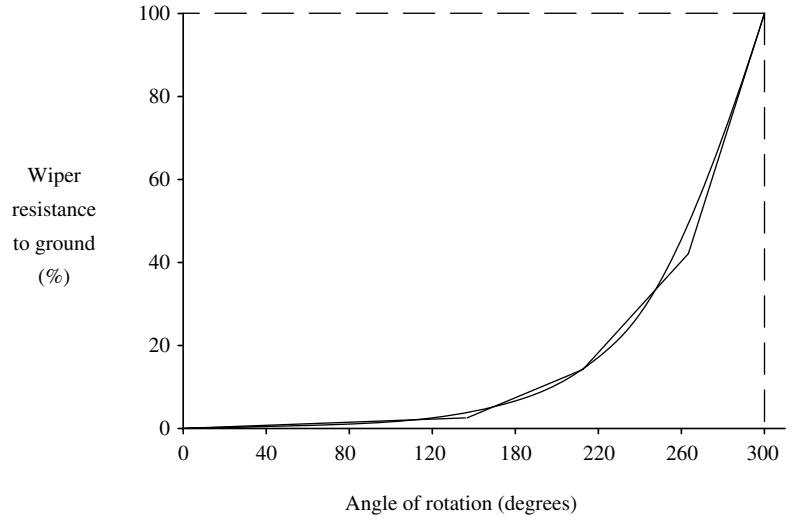


Fig. 7.7 *Approximation of logarithmic law by straight lines*

The ideal volume control would have identical attenuation (in dB) for a given amount of rotation whether that rotation was at the bottom of the range or the top.

The following QBASIC program allows you to investigate the effects of the law faking resistor. QBASIC is a version of BASIC that was bundled with the MSDOS operating system. Although recent operating systems desperately try to shake off their DOS heritage, Windows XP runs DOS programs quite happily.

Using the program, the ideal value for the law faking resistor seems to be ≈ 0.83 of the value of the potentiometer, and lower values actually make performance worse. Unfortunately, the stratagem only improves the top part of the range, but if levels can be optimized so that only the top 12 dB of attenuation are commonly used, then it works very well.

```
CLS
P = 1
PRINT ''This program calculates the step size''
PRINT ''resulting from shunting the output of a''
PRINT ''linear potentiometer with a law faking ''
PRINT ''resistor.
PRINT
PRINT ''How many steps of resolution do you want to''
PRINT ''investigate'';
```

```
INPUT N
DIM LOSSDB(N)
PRINT ''What value of potentiometer will you use'';
INPUT R
PRINT ''What value of law faking resistor will you''
PRINT ''use'';
INPUT L
PRINT
PRINT '' LOSS''; TAB(15); '' STEP''
DO UNTIL P = N
A = P * R / N
UPPER = R - A
LOWER = A * L / (A + L)
LOSSDB(P) = ((86.8589 * LOG(LOWER / (LOWER +
UPPER))) \ 1) / 10
REM THE 86.8589 FACTOR ARISES BECAUSE QBASIC USES
NATURAL LOGS
P = P + 1
LOOP
P = 2
DO UNTIL P = N + 1
CLICK = LOSSDB(P) - LOSSDB(P - 1)
PRINT ABS(LOSSDB(P)); ''dB''; TAB(15); ((10 * CLICK) \
1) / 10; ''dB''
P = P + 1
LOOP
```

A development of this idea is to provide multiple taps along the linear conductive plastic track as part of the design. *Tapped linear track attenuators* have excellent channel matching and adherence to the required logarithmic law. They often have 'Penny & Giles' written on the outside, and *feel* nice when used. It should be noted that these are not potentiometers and that measuring resistance to end stops from the wiper produces confusing results.

Switched attenuators

If quality is paramount, and we can accept a control that is not continuously variable, we could use a switched attenuator that works by selecting fixed resistors in order to control volume. The logarithmic law can now be perfect, as can channel balance.

The switched attenuator has a long and noble history. The BBC used quadrant faders (switched attenuators without detents) on sound desks until the 1970s because there is nothing worse than gently fading out a programme and hearing an abrupt change in the rate of attenuation – the ear/brain expects to hear a logarithmic decay similar to reverberation decay. Once sufficiently low noise tapped linear track attenuators became available, the (far more expensive) quadrant fader could safely be relegated to broadcasting history. Surprisingly, switched attenuators with a mere nine steps were made by Erie for consumer use in 1949,[1] but even then, the advantage of the superior law was realized.

Unfortunately, soldering resistors onto a switch to make switched attenuators takes time, but commercial attenuators for audio are once more commonly available. Switched attenuators are even available with thick film resistors fabricated directly onto the ceramic substrate of the switch wafer, and their performance is excellent.

Type 72 30-way stud switches are made by a number of UK manufacturers, and are ideal for making switched attenuators. Unfortunately, wafers may not always be interchangeable because the Winkler variant has a wider but thinner rectangular shaft compared to the Blore Edwards, Elcom, or Painton versions. Type 72 switches are about the same size as a traditional wafer switch but the contacts are silver or gold-plated brass studs precision moulded into a thermosetting plastic wafer, with a small wiper travelling across these and the centre ring. Bought new, they are expensive, but you may be able to find some in old laboratory equipment at an electronics junk shop. Haggle.

With care, they can be dismantled and their silver contacts cleaned so that they are good as new, despite being over 30 years old. To do this, carefully remove each wafer from the switch. Scrape the locking varnish off the two 8BA centre screws that hold the travelling wiper mechanism together, restrain the nuts with a spanner, and carefully unscrew them, taking care not to put stress on the wiper itself. Gently prise everything apart, taking extreme care of the plastic barrel that supports the travelling wiper, as it is very fragile. Pour a little silver polish onto clean cartridge (artist's drawing) paper, and rub the wafer, stud side down, on it until it is clean. Do the same for the wiper. Give the whole thing a squirt with electronic cleaning solvent to remove any traces of polish, and reassemble without your fingers touching the contacts. Sweat contains acid and will tarnish the contacts.

The practical disadvantage of the switched attenuator is that we can only have as many different volume levels as switch positions. Although rotary switches are available with 30 positions, as opposed to the more usual 12, this

limits us to 29 positions at best by the time that we have fitted the end stop. Commercial volume controls assume that we want an 'off' position, followed by a −60 dB position, and then provide uniform steps all the way up to 0 dB. We have already used two of our 30 positions, so 60 dB divided by 28 steps ≈2.14 dB per step. This is rather coarse, so some commercial attenuators coarsen the lower levels to allow finer control at the upper levels, but typical step size is still 2 dB.

The reason that commercial attenuators need 60 dB of attenuation range is that levels and gains are unknown. Since we are designing the pre-amplifier to match a known power amplifier, its gain can be precisely set so we can only overload the power amplifier on the last few steps of the volume control. To ensure this, we fit attenuators as necessary to individual inputs so that all incoming signals arrive at the volume control at the correct level.

In practice, we very rarely adjust the volume of a given source by more than 20 dB, so 29 switched 1 dB steps makes a surprisingly convenient volume control, although we will usually also need a mute switch.

Switched attenuator design

Assuming that we have the switches available to make a volume control, we need to be able to calculate the values of resistors required. We could do this by hand, but a computer makes life much easier. There are three basic forms of attenuator that can be used. See Fig. 7.8.

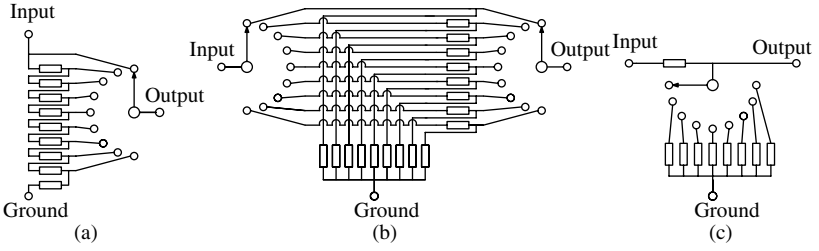


Fig. 7.8 *Basic stepped attenuators*

Figure 7.8a is similar to the carbon track attenuator in that it has a ladder of resistors from which we take the appropriate tapping, this is the form used in most commercial switched attenuators.

NB. Just because a volume control has detents, this *does not* guarantee that it is a true switched attenuator – it could be a carbon track potentiometer in masquerade. Real switched attenuators tend to be quite large. A quick test,

rather than dismantling it in the shop, which might have you thrown out, is to measure the resistance of the lower arm of each gang at the settings of near maximum attenuation with a digital multimeter. If there is any measurable difference between gangs, it is likely to be a carbon track potentiometer.

Figure 7.8b uses individual potential dividers for each setting of volume which dramatically reduces the number of soldered joints in the signal path at the expense of twice as many wafers and resistors. Additionally, the resistors must be close tolerance to maintain a good logarithmic law and channel matching.

Figure 7.8c uses one series fixed resistor, and a selection of shunt resistors in order to achieve similar performance to Fig. 7.8b at lower cost. Input resistance is no longer constant, and the series resistor must be equal to the maximum tolerable output resistance, because when this attenuator is set to maximum attenuation, its input resistance is equal to the series resistor. Thus, input resistance falls to a minimum of 25 kΩ, whereas the previous attenuators had a constant input resistance of 100 kΩ, and a maximum output resistance of 25 kΩ. This form of attenuator is best suited to driving a cathode follower because the reduced input capacitance allows the series resistor to be increased to 100 kΩ.

The following QBASIC programs generate the resistor values for the attenuators in Fig. 7.8a, b, c. They are not miracles of programming, but they are quick and easy to use, and can easily be modified for different versions of BASIC, or the key equations can be extracted and used in a spreadsheet.

The programs ask for the load resistance across the wiper, this is the grid leak resistor of the following valve. It is tempting to try to use the potentiometer as the grid leak, but this is poor practice, and can cause noise problems, it is also unnecessary, since the programs account for its loading in designing the attenuator.

This program finds values for the circuit of Fig. 7.8a. The final value given by this program is connected between the last usable switch contact and ground, it is often convenient to use one of the spare contacts on the switch as a ground terminal.

```
CLS
A = 0
B = 0
N = 0
PRINT ''This program calculates individual values''
PRINT ''of resistors between taps of the circuit''
PRINT ''in Fig. 7.8a.''
PRINT ''How many switch positions can you use'';
INPUT S
```

```
PRINT ''What step size (dB)'';
INPUT D
PRINT ''What value of resistance will be across the''
PRINT ''output of the potentiometer'';
INPUT L
PRINT ''What value of potentiometer is required'';
INPUT R
DO UNTIL N = S - 1
Y = ((R - L / 10 ^ (-A / 20)) + SQR((L / 10 ^ (-A / 20) -
R) ^ 2 + 4 * R * L)) / 2
C = R - Y - B
PRINT A; ''dB ''; C;''ohms''
B = B + C
A = A + D
N = N + 1
LOOP
PRINT A; ''dB ''; R - B;''ohms.''
```

The following program is for the circuit of Fig. 7.8b.

```
CLS
A = 0
N = 0
PRINT ''This program calculates upper (X) and lower''
PRINT ''(Y) arms of individual potential dividers''
PRINT ''for the circuit of Fig. 7.8b''
PRINT ''How many switch positions can you use'';
INPUT S
PRINT ''What step size (dB)'';
INPUT D
PRINT ''What value of resistance will be across the''
PRINT ''output of the potentiometer'';
INPUT L
PRINT ''What value of potentiometer is required'';
INPUT R
DO UNTIL N = S
Y = ((R - L / 10 ^ (-A / 20)) + SQR((L / 10 ^ (-A / 20) -
R) ^ 2 + 4 * R * L)) / 2
```

```
X = R - Y
PRINT A; ''dB ''; ''Y ='': Y; ''ohms ''; ''X ='': X;
''ohms''
A = A + D
N = N + 1
LOOP
```

The final program calculates shunt resistors for the circuit of Fig. 7.8c. Note that it *never* achieves zero attenuation, and therefore predicts the minimum unavoidable loss through the volume control and grid leak resistor (basic loss). In effect, this volume control must be considered to be a fixed attenuator plus a variable attenuator.

```
CLS
N = 0
PRINT ''This program calculates shunt resistors for''
PRINT ''the circuit of Fig. 7.8c.''
PRINT ''How many switch positions can you use'';
INPUT S
PRINT ''What step size (dB)'';
INPUT D
PRINT ''What value of resistance will be across the''
PRINT ''output of the potentiometer'';
INPUT L
PRINT ''What value of series resistor is required'';
INPUT R
B = ((-100 * LOG(L / R + L)) * 8.686) \ 1) / 100
REM THE 8.686 FACTOR ARISES BECAUSE QBASIC USES
NATURAL LOGS
PRINT ''Basic loss ='': B; ''dB, added shunt is''
PRINT ''infinite''
PRINT ''Added attenuation:''
A = B
DO UNTIL N = S - 1
A = A + D
C = R * 10 ^ (-A / 20)) / (1 - 10 ^ (-A / 20))
Y = 1 / (1 / C - 1 / L)
N = N + 1
PRINT N * D; ''dB, shunt = ''; Y; ''ohms''
LOOP
```

Spreadsheets and volume controls

As mentioned, the fundamental equations from the previous QBASIC programs can be used in a spreadsheet. The great advantage of a spreadsheet is that it can be set up to predict exact values, and then to calculate the errors that would be generated by substituting nearest standard E24 values.

We usually want to simulate a perfect 100k logarithmic potentiometer to the design of Fig. 7.8a, and load it with a 1 M grid leak resistor. Fortuitously, this combination of values produced quite low errors (ten other combinations were tried but all were worse), and a little tweaking away from obvious values improved it further – hence the 910 Ω resistor when 1k would have seemed more appropriate.

A stereo volume control of this type was built using 1% surface mount resistors and tested for error against calculated attenuation and for matching:

Loss (dB)	*R* (ideal)	*R* (practical)	Predicted error (dB)	Measured error (dB)	Stereo matching error (dB)
0	0	0	0.00	0.00	0.00
1	10068	10k	−0.01	0.00	−0.02
2	9261	9k1	−0.01	−0.03	−0.01
3	8456	8k2	0.01	−0.03	−0.01
4	7675	7k5	0.02	−0.03	−0.01
5	6932	6k8	0.02	−0.02	−0.02
6	6237	6k2	0.02	−0.04	0.00
7	5594	5k6	0.00	−0.05	0.00
8	5005	5k1	−0.04	−0.08	−0.02
9	4470	4k3	−0.02	−0.05	0.00
10	3987	3k9	−0.02	−0.05	0.00
11	3553	3k6	−0.05	−0.07	0.00
12	3164	3k0	−0.02	−0.02	−0.01
13	2816	2k7	0.01	0.00	0.00
14	2506	2k4	0.04	0.02	0.00
15	2229	2k2	0.04	0.03	0.01
16	1983	2k0	0.02	0.02	0.01
17	1764	1k8	−0.02	0.00	−0.02
18	1569	1k6	−0.06	−0.05	−0.01
19	1396	1k3	−0.01	0.01	0.00
20	1242	1k2	0.01	0.02	0.00

Table (continued)

Loss (dB)	*R* (ideal)	*R* (practical)	Predicted error (dB)	Measured error (dB)	Stereo matching error (dB)
21	1105	1k1	0.00	0.02	0.00
22	984	910 Ω	0.06	0.08	0.00
23	875	910 Ω	0.01	0.02	0.00
24	779	750 Ω	0.03	0.04	0.00
25	694	680 Ω	0.04	0.05	−0.01
26	618	620 Ω	0.02	0.02	−0.01
27	550	560 Ω	−0.01	0.00	0.00
28	490	470 Ω	0.01	0.02	0.00
	4010	3k92 + 18 Ω			

As can be seen, the predicted errors are very low, except for two instances where they rise to 0.06 dB. In practice, measured errors are a little higher, with the 8 dB step on one wafer exhibiting 0.1 dB error from the required attenuation (although this also includes test set error). Whilst 0.1 dB error would be important in test equipment, 0.1 dB deviation from perfect logarithmic law on a volume control is trivial. Even better, the worst measured matching error was only 0.02 dB.

The final value shown in the table is a resistor from the bottom of the chain down to ground, and is made up of a 3k92 0.1% and 18 Ω (any tolerance) in series. The final resistor value is critical, hence the requirement for the larger resistor to be 0.1% tolerance.

Construction

Fortuitously, the distance between the solder pins of a 'Type 72' 30 position switch can be bridged perfectly by a 1206 size surface mount resistor.

A simple jig allowing resistors to be soldered directly between the pins of the switch can be made without any special tools. The jig is simply a 12″ × 3″ piece of wood as a baseplate, with a post for gripping the switch horizontally by its shaft. The other post supports a pair of cross-lock tweezers horizontally to hold the resistor in the correct position for soldering. Precision construction is not required because the tweezers can easily be bent or rotated into the correct position. Although the tips of the tweezers may initially be too wide, they can easily be trimmed by a grinding wheel in a power drill (wear goggles). See Fig. 7.9.

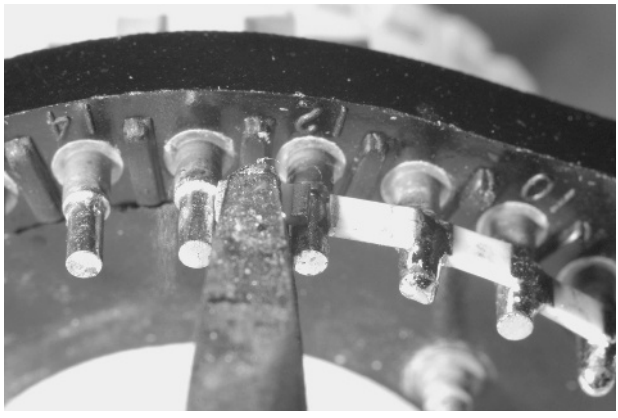


Fig. 7.9 *Soldering surface mount resistors using a jig. Note that the next soldered joint will solder one end of the new resistor and the (previously unsoldered end) of the previous resistor*

Only the end of the resistor furthest from the next resistor to be fitted is soldered at a time. The tweezers can then be squeezed to release the soldered resistor, and the switch body rotated to the next detent, perfectly positioned for the next resistor to be inserted. When the next resistor is positioned, both it and the (previously unsoldered) end of the preceding resistor are soldered. Soldering is best done by heating the switch pin first and flowing the solder onto it before completing the joint by moving the iron to heat the resistors directly. Silver loaded solder designed for surface mount components must be used.

The job is a little fiddly, and takes about half an hour per wafer – which is actually slightly faster than using wire-ended resistors. The most likely cause of trouble is losing your place on the table of values and fitting a resistor of incorrect value, so it is useful to cross resistors off the table as you place them, then double check before soldering.

Light-sensitive resistors and volume controls

Cadmium sulphide (CdS) light-sensitive resistors such as the ORP12 are occasionally mooted as possible volume control elements.

As an experiment, the author fitted a pair of ORP12 resistors into a carefully machined 2″ aluminium tube, one at each end, and added light-tight seals. A 28 V 40 mA bulb was fitted axially midway between the two resistors with a light-tight seal. One ORP12 had a dark resistance of

≈5 MΩ, and the other 100 kΩ. When fully illuminated, one had a light resistance of 69 Ω, and the other 63 Ω. Clearly, we would need to select these devices to have any hope of making a stereo volume control that tracked.

When a 1 kHz sine wave at +30 dBu was applied via a 10 kΩ series resistor, with the resistors dark, 1% distortion resulted (pure 2nd harmonic). Once the level was reduced to +8 dBu(2 V_{RMS}), the distortion fell to 0.02%. As the resistors were illuminated, both output voltage and distortion fell.

	CdS attenuator	**Switched attenuator (Fig. 7.8a)**
Distortion	Benign (2nd harmonic)	None measurable
Interchannel tracking error	Poor, needs to be selected	Easily betters 0.05 dB
Logarithmic law	Poor	Practically perfect
Ease of construction	Difficult to make light-tight	Slightly fiddly

Classic valve compressors successfully used illuminated CdS resistors as a means of controlling gain because 0.02% of pure 2nd harmonic distortion is quite a good performance for an electronically controlled gain element. A volume control doesn't need fast-responding electronic control of attenuation, so why tolerate even 0.02% distortion and poor tracking?

Balanced volume controls

It may be that you want to build a balanced pre-amplifier. In which case, the volume control should be balanced too. Curiously, some engineers, who should know better, think that this means placing identical controls in *each* path of each channel. This is wrong. The attenuators cannot be perfectly matched, so a common-mode noise signal such as hum would be attenuated unequally, converting a proportion of it to differential mode, to which the balanced pre-amplifier is sensitive.

The correct way to construct a balanced volume control, using the minimum number of components, is to use a configuration based on Fig. 7.8c. If you use the computer program to determine values, remember that the series resistor that the program uses is *twice* the value of the resistor in each leg. This form of attenuator has the disadvantage of a high output resistance when set for a sensible input resistance, and in combination with the input capacitance of the following stage, this will cause HF loss if ignored. Note that tapped track linear attenuators cannot be used as the shunt element. See Fig. 7.10.

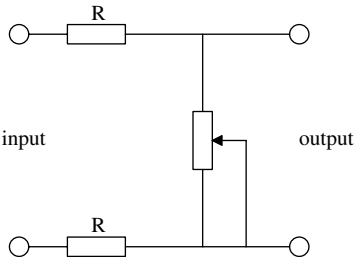


Fig. 7.10 *Balanced volume control*

Input selectors

It is highly likely that we will have a number of alternative sources to the pre-amplifier, such as vinyl, CD, digital TV or radio, etc. which need to be selected to the volume control. In this example, all the sources except vinyl are digital, so if they needed to be recorded, they would be routed elsewhere as digits. As a consequence, the output of the vinyl stage can be routed directly to the analogue record output, and the output of the recording device can be monitored using the normal input selector switch. This arrangement makes the traditional 'tape loop' switch unnecessary. See Fig. 7.11.

An unusual feature is the inclusion of a switch in the analogue record output. The vinyl stage will be designed to have low output resistance, making it perfectly capable of driving the cable capacitance to the (powered) recording device, so a buffer is not needed, but a switched-off recorder would present a non-linear load to the source in the form of unbiased transistor junctions, thus causing distortion. Most of the time, we will not be recording, so the record output can be switched off.

As the circuit stands, the rotary selector switch could suffer from crosstalk due to the capacitance between adjacent contacts (≈ 0.6 pF for a Type 72 30

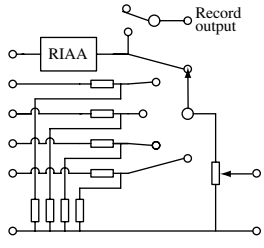


Fig. 7.11 *Input selectors with switchable record output*

position switch). If we had alternate sources plugged into the pre-amplifier, but selected an unused source, this 0.6 pF would form a high-pass filter in conjunction with the 100 kΩ input resistance of the volume control. The ear is most sensitive at ≈4 kHz, and at this frequency the *RC* combination would cause crosstalk ≈53 dB below an expected signal. On high quality traditional pre-amplifiers, this irritation was solved by having two ganged switches, one selected the source, and the other deselected the short circuit to ground on that source. Unfortunately, such wafer switches are no longer available, but a good alternative is to use a Type 72 switch with alternate contacts as inputs, and connect the unused contacts to ground, which then guards the adjacent signal contacts. See Fig. 7.12.

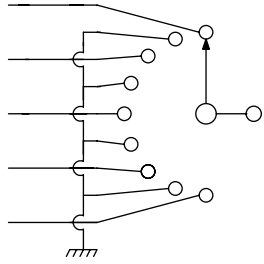


Fig. 7.12 *Reducing capacitive crosstalk on input selector by guarding*

Type 72 switches are available with two types of detent wheels. The standard type allows each stud contact to be selected (30 positions), but the coarser detent wheel selects alternate studs (15 positions). Once alternate contacts have been earthed, if you need a 'mute' position between each source, the 30 position detent is ideal, but if you want to switch directly between sources, the 15 position detent should be used.

Although crosstalk on an open circuit input could easily be at −53 dB with a traditional switch, once an input having a low source resistance is selected, the crosstalk falls rapidly. As an example, selecting a typical CD player having $r_{out} \approx 300\,\Omega$, would cause the crosstalk to fall to −107 dB. Thus, earthed alternate selector switch contacts mainly avoid nuisance crosstalk on unused sources and provide a mute switch without any extra contacts in the signal path.

Switch quality

Switches are not created equal, and suffer from a number of known defects. Stray capacitance has just been mentioned, but there are other problems.

Ideally, contact resistance should be zero, but this is never achieved. Contact resistance rises as a result of oxidation of the contacts, insufficient contact pressure and contact wear. Because of its resistance to atmospheric corrosion, gold-plated contacts are sometimes used, despite the fact that the resistivity of gold is considerably higher than that of silver or copper. It should be noted that the phenomenon of contact resistance is complex, so low-level signal switching should only be done by switches expressly designed for that purpose.

All switches have some leakage resistance, which will be specified by the manufacturer, and will be worsened by atmospheric damp.

Push-button mechanically interlocked switches should not be considered. They feel nasty to use, and are unreliable because they are generally of poor quality. Additionally, they are awkward for the amateur to use because it is difficult to cut the required rectangular hole neatly. By contrast, the rotary wafer switch only needs a drilled or punched circular hole which is then obscured.

The ideal signal switch is the mercury wetted relay. This has a droplet of mercury to 'wet' the contacts and ensures minimum, and constant, contact resistance. Because the relay has to be sealed to prevent the (poisonous) mercury from escaping, the contacts do not oxidize. The contacts do not scrape against one another, so wear is minimized.

Mercury wetted relays generally have a single closing contact, and may require correct orientation to ensure that the mercury droplet is in the correct position. If we had a series relay on each source, we could precede it with a shunt relay to ground that would normally be closed, and this would ensure almost perfect attenuation of unwanted sources. In order to protect the source from the short circuit, a 1 kΩ series resistor is normally fitted before the shunt relay, which further improves attenuation. See Fig. 7.13.

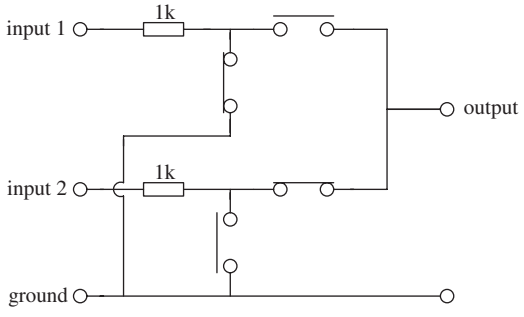


Fig. 7.13 *Shunt and series switching using relays eliminates crosstalk*

The relays should be mounted as close as possible to the input sockets, and could be connected directly to the record output relay. All of the signal switching is then as direct as possible, and the only signal wires that need to come from the rear panel go to the volume control, as the front panel selector switch only carries DC. It makes sense to choose 5 V relays, since these can then be powered by a 5 V regulator fed from the same transformer winding as the heaters. Since we are using 5 V, we could consider using logic gates to drive the relays, and use momentary action push-buttons on the front panel to select routings, or even remote control. If you choose to go down this path, choose the momentary action switches carefully, otherwise having to carefully machine identical push-buttons for each switch can be a real pain.

The *only* disadvantage to mercury wetted relays is their cost.

The RIAA disc stage

RIAA is the abbreviation for 'Recording Industry Association of America', and is the de facto worldwide post-1954 standard for equalization of microgroove records, as opposed to the numerous standards for 78s. Because the RIAA standard was not invented in Europe, but a worldwide standard *was* needed, the IEC invented an LP equalization standard that was almost identical. The only difference is that the IEC standard recommends bass cut on replay only, with a -3 dB point at 20 Hz (7950 μs) in order to reduce rumble. Most manufacturers of high quality pre-amplifiers assume that their products will be complemented by equally good turntables, and that rumble will not be a problem, so they ignore the IEC recommendation. Their equalization is therefore RIAA.

Nevertheless, there is considerable pressure to modify RIAA stages to include a low frequency roll-off because:

- Some valve power amplifiers are susceptible to output transformer core saturation if high amplitude signals are applied at low frequencies (< 50 Hz).
- Bass reflex loudspeakers are easily overloaded at frequencies below their acoustic roll-off because there is negligible damping of cone motion. Bookshelf reflex loudspeakers tend to roll off below 100 Hz, whereas freestanding reflex loudspeakers could improve this to 50 Hz, or less.
- Vinyl records contain low frequency (< 20 Hz) noise due to warps and rumble.

It is therefore argued that these problems could be avoided by implementing some form of LF roll-off within the RIAA stage. One possibility is to

implement the IEC 7950 μs recommendation, but a more sophisticated approach is to incorporate a properly designed high-pass filter having a final slope of 12 dB/octave, or more, set at ≈10 Hz.

The author firmly believes that neither of the preceding electrical approaches is correct, and that RIAA equalization should be reserved solely for correcting the record equalization applied by the manufacturer at the time of cutting. CD players do not add a 10 Hz high-pass filter to solve the problems of poorly designed loudspeakers or questionable output transformers, so why adulterate vinyl? Warps and rumble are mechanical problems, and should therefore have mechanical solutions, not electrical 'fixes'.

The mechanical problem

Fortunately, a 12 dB/octave high-pass mechanical filter is unavoidably formed by the compliance of the cartridge suspension and the effective mass of the arm plus cartridge. The low frequency arm/cartridge resonance may be found using the standard resonance equation:

$$f = \frac{1}{2\pi\sqrt{C m_{\text{total}}}}$$

where: C = cartridge vertical compliance
m_{total} = total effective mass

Typical values might be:

Cartridge mass:	5 g
Mounting hardware (screws and nuts):	1.5 g
Arm effective mass:	12 g
Total mass (m_{total}):	18.5 g
Cartridge vertical compliance (C):	15×10^{-6} dyne/cm

The previous figures applied to a unipivot arm designed for an Ortofon Quattro moving coil cartridge with its outer body removed, and resulted in the ideal resonant frequency of 10 Hz.

It has been suggested that a higher frequency (12–15 Hz) should be set, as this will be more effective in reducing low frequency noise, and this is quite true. However, we live in a practical world and dramatically reducing an arm's effective mass (to raise the resonant frequency) inevitably produces a flimsy structure only suitable for cartridges that do not transfer much vibration into the arm. Unfortunately, such cartridges are high compliance, and we are

back where we started. Additionally, even setting the resonant frequency as low as 10 Hz means that the reproduced response (when RIAA equalized) is likely to be −1 dB at 20 Hz, depending on damping.

As an aside, the best way of reducing effective mass is to remove mass at the headshell. Modern arms are usually fixed headshell, so this leaves the cartridge. A moving coil cartridge often has a heavy outer shell that can save valuable grammes *if* it can be removed without damaging the internal workings. Even better, provided that the magnet system is firmly anchored, removing the body eliminates box resonances. The stylus cantilever is now completely exposed, which makes cueing and alignment easy, but leaves it frighteningly vulnerable to damage. All factors considered, the trade is well worthwhile, which is why some cartridges are sold naked.

Even if the arm/cartridge resonant frequency is correct, the mechanical high-pass filter will only operate correctly if the resonance is correctly damped. The general principle is that the moving pick-up arm is fitted with a paddle which is forced to move through a viscous liquid, thus damping the motion of the arm. Ideally, the damping should be applied at the headshell because this greatly reduces the energy transferred from the cartridge into the arm, and particularly reduces excitation of unavoidable high frequency structural resonances, but damping near the pivot, as required by almost all unipivots, damps the low frequency resonance equally well.

Mechanical damping has to be set by trial and error, and commonly, far too much damping is applied – the fluid is either too viscous, or there is too much of it. One way to set damping is to play a badly warped record with no damping applied, and observe cartridge movement as the warps are traversed. If the cartridge appears to bounce relative to the record surface, add a little fluid, and try again. Use as little damping as possible, as too much will *increase* low frequency noise and cause tracking problems at higher frequencies as undistorted stylus displacement is squandered by tracking warps rather than recorded signals.

Setting the mechanical high-pass filter correctly has two major consequences. First, it means that we no longer need an electrical high-pass filter, but more importantly, it means that stylus vertical deflection is considerably reduced and distortion generated by the cartridge falls.

Arm wiring and moving coil cartridge DC resistance

Although universally ignored, the wiring resistance of a pick-up arm can become significant, particularly when a low output moving coil cartridge is

stepped up by a transformer, because the transient response of the associated transformer is critically dependent on source resistance.

The author measured the resistance of a 5 m length of the fine wire used within his pick-up arm, and found that it had a resistance of 0.45 Ω/m, so a typical 9″ arm requiring 600 mm of wire for each channel (loop to cartridge and back) contributes 0.27 Ω resistance. The 1 m loop of 0.7 mm silver twisted pair from the arm base to the pre-amplifier added 0.12 Ω, bringing the total arm resistance to 0.39 Ω. Pick-up arm wiring modifications are popular, and in order to reduce the number of joints, some take the fine wire required within the arm tube all the way to the pre-amplifier input plug. 600 mm leads are typical, so this increases the loop resistance to 0.81 Ω. To put these values into perspective, the previously mentioned Ortofon Quattro cartridge had a specified coil resistance of 3 Ω.

High levels at high frequencies imply high acceleration of the stylus tip. In order to minimize the force required by the vinyl to accelerate the tip ($F = ma$), cartridge manufacturers strive to minimize stylus tip mass. The most effective way of reducing tip mass is to use a smaller diamond, because very little of the diamond actually constitutes the stylus tip, but this makes the diamond harder to grip whilst grinding and lapping. However, the mass of the coils in a moving coil cartridge is also significant, so a useful improvement can be obtained by reducing wire diameter. Unfortunately, cartridge manufacturers do not always update their technical information as they develop a product, so the measured DC resistance of the coils can be higher than specified. The author recently measured a nominal 6 Ω cartridge and found its DC resistance to be 10.5 Ω.

The ideal point to measure cartridge DC resistance is at the pre-amplifier input plug because this takes arm wiring resistance into account. The fine wire within a cartridge could be ruptured by passing a high current through it, so the author measured the current supplied by his meter when set to the lowest resistance range, and found it to be 0.1 mA – quite low enough not to disturb a cartridge. If in doubt, check the specification for your meter, and bear in mind that maximum current will be sourced by the lowest range (which is the one you will use).

Once the source resistance seen by the step-up transformer is accurately known, its optimum output loading can be found. (See Chapter 4.)

Designing the RIAA stage

If power amplifiers are let down by their phase splitters, then the Achilles' heel of the pre-amplifier must surely be the RIAA disc stage. The stage has

to satisfy so many contradictory requirements simultaneously that its design and execution are fraught with problems.

When we investigated power amplifiers, we looked at some classic designs to see how the goals were achieved. There were *no* classic RIAA stages, they varied from mediocre to plain awful.

This was not always due to incompetence on the part of the designers. They had poorer quality components, and could not use regulated HT supplies as is habitual today. However, the main factor was that there was simply no incentive to design superb disc stages, because the signal leaving the turntable was not very good. Vinyl was regarded as a very poor quality source, often requiring low-pass filtering at 8 kHz to reduce dust/click disturbance. Amazingly, good vinyl turntables and cartridges were all available, but the appalling mechanical/acoustical failings of most of the arms and chassis/plinths meant that the electronics engineers were never exposed to criticism.

Determination of requirements

As for the line stage, we need to define the detailed requirements of the RIAA stage before we can begin the design.

1 *Low noise and no hum*. We have to admit that valves are not as quiet as the latest generation of low noise IC op-amps, but DC heater supplies eliminate hum and slightly reduce valve noise. Pentodes are complete non-runners, and we will need to be careful in our use of triodes.

2 *Constant input resistance and capacitance*. This might seem obvious, but many designs fail to appreciate that moving magnet cartridges are particularly sensitive to changes in electrical loading.

3 *Accurate RIAA*. It is quite unbelievable how many designs have incorrect RIAA equalization (ancient and modern). This is either down to a failure to use the correct equations, or to appreciate the loading conditions.

4 *Low sensitivity to component variation*. Valves wear out, and as they do so, r_a rises. Similarly, when a valve is replaced, the new value of C_{ag} may not be the same as the old valve. Neither of these effects should noticeably affect the accuracy of RIAA equalization.

5 *Good overload capability*. But what capability is necessary? Using a Tektronix TDS420 digitizing oscilloscope, the maximum output of LPs was investigated in conjunction with a high quality record playing system. The TDS420 was first used in 'envelope' mode to find the

maximum output of the cartridge, and monitored an entire day of listening to music. The largest musical peaks were found whilst playing a Mobile Fidelity pressing of Beethoven's 9th Symphony. Before equalization these peaks rose to +16 dB above the nominal 5 cm/s level, but clicks due to dust or scratches rose to about twice this level at +22 dB. See Fig. 7.14.

Individual clicks were then captured, and it was found that the vinyl/tip mass resonance was being excited, and that this produced a heavily damped oscillation at 56 kHz for this particular (moving coil) cartridge. See Fig. 7.15.

Ultrasonic overload could either generate intermodulation products that come back down into the audio band, or worse, it could cause blocking. Blocking is particularly undesirable because it converts a momentary overload that might have been almost unnoticeable into a protracted low frequency disturbance whose severity is amplified by RIAA equalization. If a power amplifier blocks, the user can turn the volume down, but this is not possible in an RIAA stage, so blocking must be avoided at all costs.

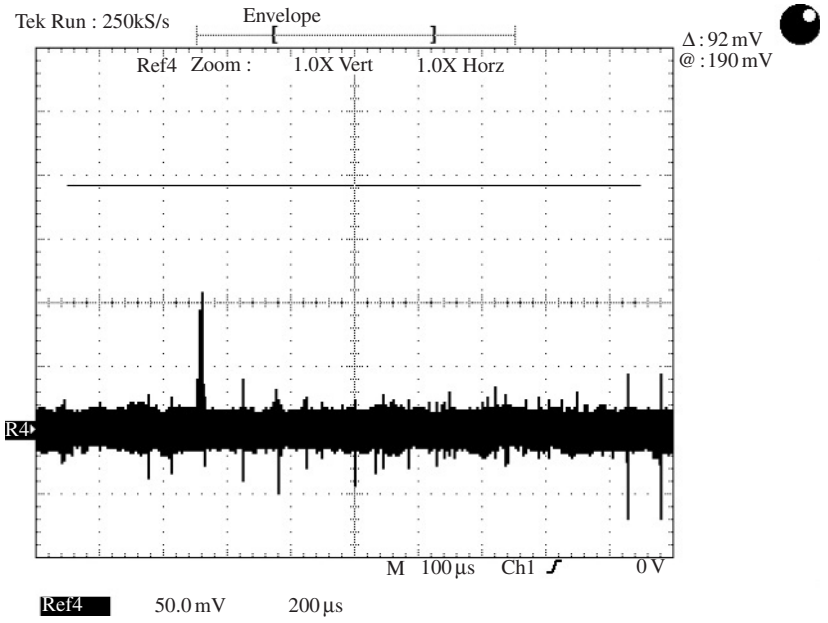


Fig. 7.14 *Unequalized enveloped music output from cartridge (peaks are dust/clicks)*

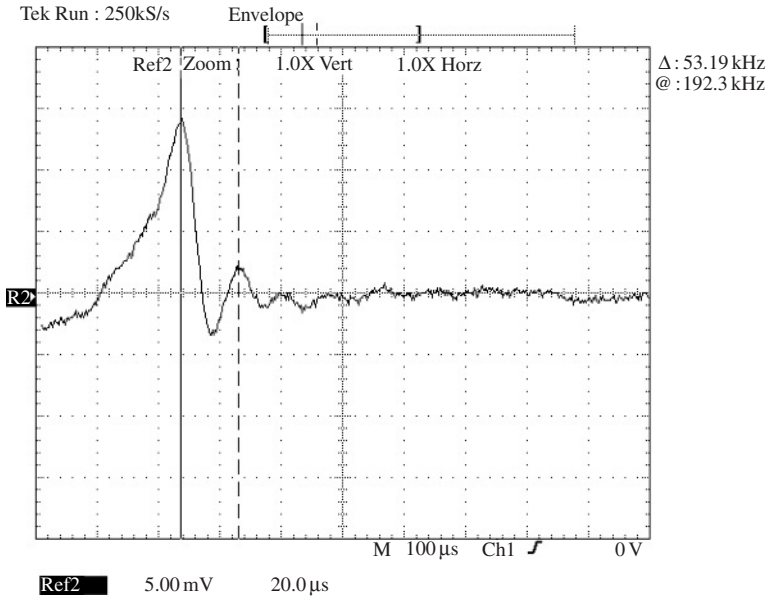


Fig. 7.15 *Unequalized output from cartridge showing excitation of vinyl/tip mass resonance*

We should now allow for variable cartridge sensitivity of about 6 dB – if we need more than this, we should reconfigure the RIAA stage.

A good design should not operate permanently at its limits, so a further 6 dB margin is desirable, to give a total headroom of 28 dB in the audio band, *rising* to 34 dB or more at ultrasonic frequencies. Very few pre-amplifiers of any age achieve this requirement and low noise simultaneously.

Worn/old discs generate more ultrasonic energy than a new disc. This may be due to dirt ground into their groove, or because they were played by a cartridge that mistracked, causing wall damage as the stylus flailed helplessly from side to side of the groove. Inadequate ultrasonic overload margin is the reason why a poor pre-amplifier can make a worn record sound unplayable, but a good pre-amplifier extracts the best from the disc.

6 *Low distortion*. This is an obvious requirement, and is linked to (5).

7 *Low output resistance*. The RIAA stage should be able to drive cable capacitance, either to a recording device, or to the line stage if it is remote from the RIAA stage.

8 *Low microphony*. Valves are always microphonic, but it is possible to make them worse. A low value of anode load combined with high anode current reduces noise, but increases the *power* gain of the stage, which causes microphony to rise. Ultimately, we have to isolate the valves mechanically. This is less of a problem than it seems, since most of the structural resonances of the electrodes are above 1 kHz, and a mechanical filter to deal with this is easily made by floating the RIAA stage sub-chassis on springs of knicker elastic. (Metal springs tend to ring at the same frequencies as the internal structures of valves.)

Now that we know our requirements, we can consider which topology would suit.

The low noise and constant input resistance requirements eliminate shunt feedback. Low noise rules out the pentode. We are therefore left with a combination of triode stages having active equalization determined by series feedback, or with passive equalization. Either of these contenders may be further broken down into performing the equalization 'all in one go', or splitting it over a number of stages.

RIAA equalization

To see how we need to tackle the problem of RIAA equalization, we need to define RIAA equalization. The equalization is specified in terms of time constants: 3180 μs, 318 μs, and 75 μs. The equation that generates the gain G_s required by the RIAA replay equalization standard is:

$$G_s = \frac{(1 + 318 \times 10^{-6} \times s)}{(1 + 3.18 \times 10^{-3} \times s)(1 + 75 \times 10^{-6} \times s)}$$

where $s = j\omega$ and $\omega = 2\pi f$.

This is not a friendly equation, and a computer program is the easiest way of subduing it. Running the following computer program yields the data shown in the table, and calculates the required response of the equalizer *only*, a perfectly pre-equalized signal passed through a perfect equalizer would yield an amplitude response deviation of 0 dB, and a phase response of 0° for all frequencies.

```
CLS
OPTION BASE 1
DATA
0,10,20,50.05,70,100,200,500.5,700,1000,
2000,2122,5000,7000,10000,20000,50000,70000,
100000,200000
FOR R = 1 TO 20
READ F
W = 2 * 3.1415927# * F
A =.000318
B =.00318
C =.000075
REALU = 1 - W ^ 2 * B * C + W ^ 2 * A * (B + C)
IMAGU = W * (A - B - C - W ^ 2 * A * B * C)
LOWER = (1 - W ^ 2 * B * C) ^ 2 + W ^ 2 * (B + C) ^ 2
MAG = SQR(REALU ^ 2 + IMAGU ^ 2) / LOWER
GAIN = (((19.91102 + 8.68589 * LOG(MAG)) * 1000) \ 1) /
1000
REM THE 8.686 FACTOR ARISES BECAUSE QBASIC USES NAT-
URAL LOGS
PHASE = ((572.96 * ATN(IMAGU / REALU)) \ 1) / 10
PRINT F, GAIN, PHASE
NEXT R
```

Note that correct syntax is vital, and the numbers following the DATA statement must all be typed in on one line even though they cannot be fitted onto a single screen (or page) width.

Note that although the amplitude response has been normalized to a 1 kHz reference, the phase response has been left untreated.

The table shows that considerable gain is needed at low frequencies, whilst HF attenuation must continue indefinitely, which excludes the series feedback 'all in one go' topology, because its gain cannot fall below unity. Although this failing can be exactly compensated after the feedback amplifier, it does mean that the response before compensation is rising, which compromises distortion and ultrasonic headroom within the amplifier.

Frequency (Hz)	Gain (dB ref. 1 kHz)	Phase (degrees°)
0	19.911	0
10	19.743	−10.4
20	19.274	−20
50.05	16.941	−40.6
70	15.283	−48.4
100	13.088	−54.8
200	8.219	−59.6
500.5	2.643	−52.6
700	1.234	−49.7
1 000	0	−49
2 000	−2.589	−55.9
2 122	−2.866	−56.9
5 000	−8.210	−72.1
7 000	−10.816	−76.8
10 000	−13.734	−80.6
20 000	−19.620	−85.2
50 000	−27.541	−88.1
70 000	−30.460	−88.6
100 000	−33.556	−89
200 000	−39.575	−89.5

'All in one go' equalization

Because the 1 kHz level is ≈ 20 dB below the maximum level at LF, any 'all in one go' passive network *must* have ≥ 20 dB of loss, since the network is in parallel with the grid leak resistor of the following valve, which causes additional attenuation. It is difficult to design a pre-amplifier of acceptable noise and overload capability using this network, so this topology can usually be excluded.

If we should decide to use either of the two previous topologies, the relevant formulae are given in the definitive AES paper by Lipshitz.[2]

Of the four possible networks that Lipshitz gives, these reduce to two for passive equalization. Of these two, only one has a capacitor in parallel with the lower arm of the network. This feature is important because it allows us to account for stray and Miller capacitance, so it is therefore the only feasible network in a valve pre-amplifier. See Fig. 7.16.

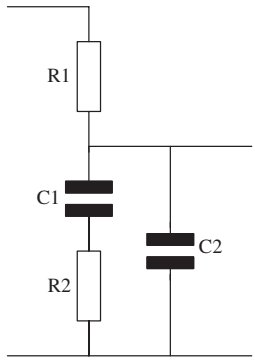


Fig. 7.16 *Passive RIAA de-emphasis network*

The relevant equations for this *passive* network are:

$$R_1C_1 = 2187\,\mu s$$
$$R_1C_2 = 750\,\mu s$$
$$R_2C_1 = 318\,\mu s$$
$$C_1/C_2 = 2.916$$

These numbers are exact, and have not been rounded.

Remember that any grid leak resistor in parallel with the lower arm of the network, or non-zero output resistance of the driving stage changes the effective value of R_1 as seen by the network. Therefore, the values for the network must be calculated using the Thévenin resistance seen by that network.

Likewise, any stray, or Miller, capacitance must be subtracted from the calculated value of C_2.

For any 'all in one go' topology other than the above network, it is essential to refer to the Lipshitz paper, and read it thoroughly before embarking on design.

Split RIAA equalization

We are now left with only two possibilities for equalization, split active and split passive, so we must define how to split the equalization. Fortunately, there is only one rational way to split the equalization, and that is to pair the 3180 μs with the 318 μs, but to perform the 75 μs separately.

The 75 μs time constant defines a low-pass filter whose $f_{-3\,dB} \approx 2122$ Hz and rolls off at 6 dB/octave thereafter. This would be an ideal filter to use

early in the pre-amplifier since it would allow HF overload capability after that stage to rise at 6 dB/octave above cut-off, which is *exactly* what we need.

The 75 μs time constant can be implemented passively following the input stage, which has the advantage of ensuring that the load seen by the cartridge is constant with frequency.

Moving magnet cartridges often use the load capacitance in conjunction with the generator's self-inductance to form a resonant equalizer that corrects the falling mechanical response of the cartridge. Thus, the value of load capacitance is critical, but this can be set quickly and easily by adding a twin-gang ≈300 pF air-spaced variable capacitor salvaged from a (probably valve) medium wave radio. See Fig. 7.17.

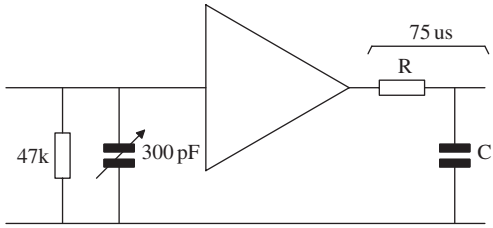


Fig. 7.17 *Catridge loading and 75μs*

The main reason for the choice of a passive 75 μs RIAA equalization network is that a series feedback amplifier cannot make $A_v < 1$, and a shunt feedback amplifier would have noise problems. Additionally, although it was not noted earlier, the output stage of a feedback amplifier attempting this response faces a heavy capacitive load. Briefly, the capacitive load demands a large current at HF, and is equivalent to changing the AC loadline to a far lower value of load resistance, which results in additional distortion *before* closing the feedback loop.

It should be noted that *all* of the observations so far are equally relevant to discrete semiconductor or IC-based pre-amplifiers.

Split RIAA equalization for valves

The 3180 μs, 318 μs pairing defines a shelf response with a level variation of exactly 20 dB. Using IC op-amps it is equally convenient to perform this actively or passively, but with valves it is more convenient to use passive equalization.

We have now defined the optimum way of achieving RIAA equalization in a valve pre-amplifier. We should use passive 75μs, followed by passive paired

3180 μs, 318 μs and we should amplify using triode valves. These decisions enable us to draw a block diagram of the pre-amplifier. See Fig. 7.18.

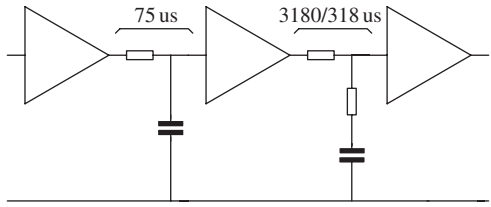


Fig. 7.18 *Block diagram of RIAA pre-amplifier*

Drawing the block diagram was useful, because it has further defined the design by requiring a minimum of three stages. Note that the block diagram has completely ignored practicalities such as coupling, or decoupling, capacitors and grid leak resistors. Nevertheless, it represents a simplicity of design to which we should aspire, i.e. DC coupling throughout. This ideal is achievable, but it is not ideal for the novice constructor, so we will be a little more cautious in our first design.

Noise and input capacitance of the input stage

Bearing in mind our previous requirements for low noise, the first stage is the *crucial* stage, and must have low noise above almost every other requirement. This is reasonable, because even +34 dB ref. 5 mV is only 700 mV_{pk-pk}, so linearity ought to be a minor problem.

Designing for low noise usually means wringing the utmost gain out of the first stage such that noise considerations in succeeding stages are irrelevant. This would imply using a high μ triode such as the ECC83 or ECC808 (electrically almost identical, but lower hum and noise, and completely different pin-out), but with a typical gain $A_v = 70$, this would result in an input capacitance of ≈120 pF including strays.

Most moving magnet cartridges are designed to be loaded by a specific capacitance, and older Shures and Ortofons needed 400–500 pF, but more modern cartridges tend to need 250 pF. Once we include pick-up arm wiring capacitance and connecting cable capacitance to the 120 pF contributed by the ECC83, the loading capacitance seen by the cartridge could rise to 300 pF. The ECC83 is probably now out of the running, unless we are prepared to

rewire the arm (which might not be such a bad idea), and we are back to the E88CC with a lower gain and lower shunt capacitance; even the *SN7/*N7 is usable provided that we take a little care over the noise of the next stage.

Although high μ Loctal valves such as the 7F7 are a possibility, high μ octal-based triodes are almost certainly forbidden because of their excessive C_{ag}. The 6SL7GT, which with $\mu = 70$, is the predecessor of the ECC83, has $C_{ag} \approx 2.8\,pF$ (RCA value: depends on manufacturer/source). With a typical gain of 50, this would result in an input capacitance, including strays, of 160 pF. This is now uncomfortably close to our 250 pF limit, and once arm wiring capacitance is included, could only be achieved by mounting the entire RIAA stage directly below the pick-up arm mounting so that the internal wires of the arm connected directly to the grid.

Mounting the RIAA stage onto the plinth, directly below the arm mounting, has enormous advantages in terms of input capacitance, rejection of induced noise and microphony. It also makes the turntable completely non-standard, and may not even be physically possible, due to limited space, or limited weight carrying ability. A delicately suspended sub-chassis turntable will not take kindly to a pound or two (500–1000 g) of pre-amplifier hanging from the arm mounting. Conversely, a turntable, such as the Garrard 301, that *must* be directly mounted onto a very heavy plinth would scarcely notice the extra mass.

The E88CC has an additional advantage in that r_a is low, and, as we will soon see, this helps noise performance. Additionally, a low r_a forms a small proportion of the total resistance that defines the 75 μs roll-off, which then satisfies our earlier requirement of reduced sensitivity to component ageing and changes.

Noise in the input stage is not only determined by the valve, but also by the associated resistors, of which R_L is by far the most important. See Fig. 7.19a.

In order to be able to calculate the noise performance of the stage, we need to redraw the circuit as a simple equivalent circuit, which makes analysis easier. See Fig. 7.19b.

We have replaced the output of the valve with a perfect Thévenin voltage source, and r_a has been included. A moving magnet cartridge can be represented as a resistor in series with an inductor, and since a Thévenin source has zero resistance, we could replace it with a short circuit, and redraw the circuit yet again. See Fig. 7.19c.

We are now in a position to add some noise sources to our equivalent circuit. See Fig. 7.19d.

The derivation of this final equivalent circuit was taken in many steps because the final circuit bears very little resemblance to the original circuit.

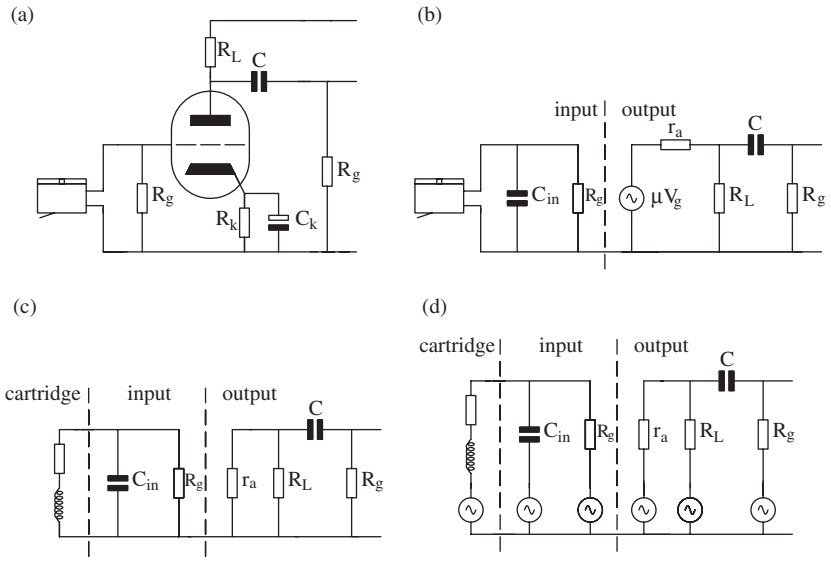


Fig. 7.19 *Noise in the input stage*

Before embarking on complex calculations, we can make some important, and useful, observations.

All of the noise sources (with their associated resistances) after the valve are in parallel, so a source of zero resistance will short circuit *any* other source, provided that there is no additional series resistance. Modern designs aim for $R_g \approx 100r_a$, and $R_L \approx 10r_a$, so r_a tends to shunt these other sources. This should make the contribution of R_g insignificant, so that any convenient value of R_g could be used, but the series coupling capacitor reduces the shunting effect of r_a. The reactance of this capacitor is:

$$X_c = \frac{1}{2\pi fC}$$

For a typical grid leak of 1 MΩ, we might use a coupling capacitor of 100 nF to give a −3 dB frequency of 1.6 Hz. If we assume that the lowest noise frequency of interest is 20 Hz (and this is debatable), then we find that at 20 Hz, $X_c = 80\,k\Omega$. This is such a high value that it nullifies any possible shunting effect by r_a, until X_c falls to a value lower than r_a.

The result of this is that the usual choice of coupling capacitor does *not* allow r_a to shunt the noise from the grid leak resistor at frequencies below 1 kHz. The resistor therefore produces noise whose amplitude is inversely

proportional to frequency (1/*f* noise), but that rises to the maximum theoretical thermal noise for that value of resistor ($v_n = \sqrt{4kTB}$).

To prevent this excess noise, we might decide to use a value of coupling capacitor sufficiently large that r_a is able to shunt R_g at *all* frequencies, which would require a value ≈10 μF. This is a large capacitor, and DC coupling is preferable if possible, but the technique has been used in a number of commercial pre-amplifiers.

Assuming that we have dealt with the grid leak resistor and the coupling capacitor, we are left with the anode load resistor R_L, and the valve itself, which leaves us with a simple equivalent circuit. See Fig. 7.20.

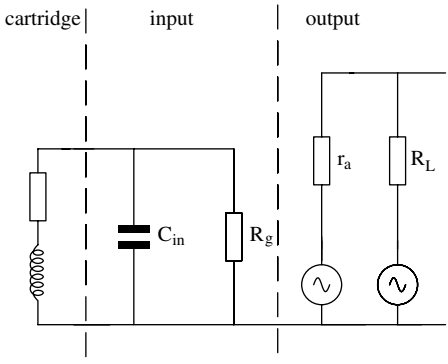


Fig. 7.20 *Final equivalent circuit for noise sources in the input stage*

R_L generates thermal noise, and if it is a film resistor, it also generates *excess* noise. Excess noise is generally specified by manufacturers in terms of μV/V of applied DC. We will therefore investigate a typical stage. See Fig. 7.21.

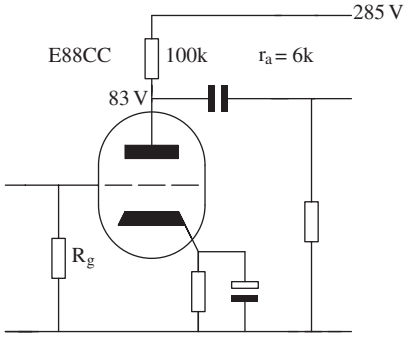


Fig. 7.21 *Typical input stage for noise analysis*

The DC voltage across $R_L \approx 200\,V$. A typical 100 kΩ 2 W metal film resistor generates 0.1 μV/V of excess noise, so 20 μV would be generated in this circuit. The thermal noise of a resistor is given by:

$$v_n = \sqrt{4kTBR}$$

where: k = Boltzmann's constant $\approx 1.381 \times 10^{-23}$ J/K
T = absolute temperature in K = °C + 273.16
B = bandwidth in Hz
R = resistance in Ω

For a typical internal temperature of 40 °C (313 K), with a bandwidth of 20 kHz, this is more conveniently expressed as:

$$v_n = 1.86 \times 10^{-8} \sqrt{R}$$

Using this equation, we find that a perfect 100 kΩ resistor generates 5.9 μV of thermal noise. In this instance, the resistor's thermal noise has been greatly exceeded by its excess noise. To find the total noise of the resistor, we must add the individual noise *powers*, which, if we remember that $P = V^2/R$, means that:

$$v_{\text{noise (total)}} = \sqrt{v_1^2 + v_2^2 + v_n^2 \ldots}$$

This gives a total noise for the resistor of 21μV, and was rather tedious, but it demonstrates two points.

- For wirewound resistors we need only calculate the thermal noise. (No excess noise.)
- For metal film resistors we need only calculate the excess noise. (This simplification works because in practical circuits, as the DC voltage across the resistor falls, so does its value.)

Now that we have simplified the noise sources in the resistor, we can see how they will be shunted by the r_a of the valve, and redraw the circuit. See Fig. 7.22.

It is now easy to see that the circuit is a potential divider, and that the actual contribution of resistor noise to the circuit is equal to the open circuit resistor noise multiplied by the attenuation of the potential divider. In our example, this reduces the noise of the resistor from 21 μV to 1.26 μV. It

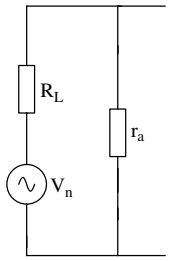


Fig. 7.22 *Effect of r_a on noise produced by R_L*

should be noted that if R_k is left unbypassed, r_a rises dramatically and is no longer able to shunt resistor noise.

If we divide the noise voltage by the gain of the stage $A_v = 29$, we can find the *input referred noise*, which is 43 nV. The significance of this is that it enables us to sum this noise with any noise sources at the grid, such as the grid leak resistor. In practice, if we calculate the thermal noise generated by R_g and its attenuation by the cartridge, we generally find that it is insignificant compared to the valve noise. In any event, we do not have a choice about R_g since it is set by the cartridge.

Valve noise

We should now consider the noise generated by the valve itself. Valves produce noise because I_a is made up of individual electrons that shower the anode, and also because electrons leave the cathode with random velocities to join the space charge, so this implies that cathode chemistry could affect noise.

There is a handy little approximation used by RF engineers:

For triodes: $$r_{eq.} \approx \frac{2.5}{gm}$$

This says that the *white* noise generated by the valve is equivalent to the thermal (white) noise generated by a perfect resistor r_{eq} at the *input* of the valve. For our example triode, $gm \approx 5.3$ mA/V, so the equivalent noise resistance would be 470 Ω.

Using $v_n = 1.86 \times 10^{-8} \sqrt{R}$, the input voltage noise produced by the valve is therefore ≈400 nV and swamps the 43 nV (input referred) noise produced by the anode load resistor (as it should in a good design), and we need not sum the noise powers of the valve and the resistor.

For pentodes[3]: $$r_{eq.} \approx \frac{I_a}{I_a + I_{g2}} . \left(\frac{2.5}{gm} + \frac{20 I_{g2}}{gm^2} \right)$$

Applying this equation to the low noise EF86 pentode operating at $I_a = 1.25\,mA$, $I_{g2} = 0.3\,mA$ predicts a noise resistance of $3.9\,k\Omega$, and a noise voltage (20 kHz bandwidth) of $1.2\,\mu V$. However, Mullard measured $2\,\mu V$ for a noise bandwidth of 25 Hz–10 kHz under the same DC conditions, which corresponds to $2.8\,\mu V$ for a 20 kHz bandwidth.

Flicker noise

The preceding noise equations are of limited practical use at audio frequencies because they do not account for $1/f$ noise, or *flicker* noise, but they do indicate that pentodes are much noisier than triodes, and that *gm* should be maximized. Unfortunately, there is no way of predicting flicker noise for a valve as it is highly sample dependent, although its level is directly related to the level of thermal noise for a given valve. Valves having pure tungsten cathodes do not produce flicker noise – but they are hopelessly microphonic, and their *gm* is low.

Alternative input valves

Although audio frequency noise predictions have been dismissed as being inaccurate in an absolute sense, if we already had a pre-amplifier whose noise performance was deemed acceptable, we could calculate its noise performance and compare it to the predicted noise performance of a new design. This would be extremely useful if the existing cartridge had a worn stylus and a new cartridge of differing sensitivity was being contemplated. Since noise is generally more of a problem for moving coil cartridges, the possibilities for the new cartridge are:

Cartridge output voltage	**Valve choice considerations**
Significantly reduced (−6 dB)	Transformer essential: noise questionable
Identical (0 dB)	Choice: retain transformer and improve noise, or trade noise against transformer problems
Significantly increased (+6 dB)	Transformer questionable: a quiet valve could make it unnecessary

The assumed dominant noise source is the input valve, so its *gm* at its operating point must be compared with proposed alternatives. Relative noise can be found using:

$$N_{(\mathrm{dB})\ \mathrm{relative}} = 10 \log \frac{gm_{\mathrm{current\ valve}}}{gm_{\mathrm{proposed\ valve}}}$$

Example

A pre-amplifier with an input stage having a *gm* of 5.3 mA/V was initially designed to be used with a moving coil cartridge plus 1:10 step-up transformer to raise the input to the pre-amplifier to 2 mV_{RMS} @ 5 cm/s. A possible replacement cartridge is claimed to produce >500 μV at the same recorded velocity, so it might be possible to dispense with the input transformer. The Loctal STC 3A/167M and Magnoval WE 437A single triodes were designed for telephone repeater amplifiers, and operated with 40 mA of anode current have a *gm* of 47 mA/V, so:

$$N_{(\mathrm{dB})\ \mathrm{relative}} = 10 \log \frac{5.3}{47} = -9.5\,\mathrm{dB}$$

The proposed input valves are 9.5 dB quieter, but the proposed cartridge is 12 dB lower output, so the signal to noise ratio would be degraded by 2.5 dB. In practice, the step-up transformer would have introduced some loss, 1 dB is typical, so the true deterioration in signal to noise ratio is likely to be around 1.5 dB – hardly noticeable. Unfortunately, the proposed valves are virtually unobtainable in usable condition, but a triode connected NOS E810F easily achieves a *gm* of 50 mA/V, and is readily available, so this is a practical possibility.

Alternatively, we can increase *gm* by connecting a number of devices in parallel, since the noise falls by a factor of $\sqrt{n}$. The LM394 super-match transistor is an extreme example of this technique, as it contains a pair of composite transistors each made of 100 individual devices to give a 20 dB improvement. Paralleling 100 E88CC is impractical, but a worthwhile, if somewhat modest, improvement of 4.5 dB can be gained by using three devices in parallel. Note that the input capacitance trebles, outlawing input transformers, moving magnet cartridges, and even some moving coil cartridges. (The high output moving coil Sumiko 'Blue Point Special' specifies maximum load capacitance as 200 pF.)

Unfortunately, the previous examples demonstrate an important point. Whilst we may improve noise by a better choice of input valve, or valves, we pay dearly for quite small improvements, since obtaining a high *gm* is expensive and invariably current hungry. To minimize noise, it is always better to present the input stage with a healthy signal, rather than hope to amplify a weak one cleanly.

Noise improvement due to RIAA equalization

RIAA equalization reduces the noise equivalent bandwidth[4] to 118 Hz, which improves noise by 22.3 dB, but because the equalization requires a gain of 19.9 dB referred to 1 kHz, the final improvement due to equalization when measuring a pre-amplifier is a meagre 2.4 dB.

Valve noise summary

Despite all the previous caveats, qualifications and provisos, we *can* make some useful generalizations that will avoid unnecessary calculations when designing for low noise:

- Pentodes are significantly noisier than triodes.
- Sample variation of both types can be large. (Flicker noise is largely determined by the cleanliness of the room in which the valve was assembled, so although a given manufacturer tends to be consistent, there can be differences between manufacturers – or more accurately, their assembly rooms.)
- To render the noise of R_L insignificant, there must be no feedback at the cathode, since this reduces the shunting effect of r_a. This is also true for a μ-follower, even though omitting C_k has no discernible effect on gain. The cascode has $r_a \approx \infty$, so the noise from R_L must be considered.
- Maximize *gm* for low noise, either with a single excellent valve, or with a number of lesser valves in parallel. (This maxim holds true whatever the device.)
- Maximized *gm* invariably raises the input capacitance of the input stage and usually precludes moving coil step-up transformers.
- Excess noise dominates in film resistors passing DC. Wirewound and bulk foil resistors do not produce excess noise.
- A very large (typically 100 times normal) coupling capacitor allows r_a to shunt the noise generated by the grid leak resistor of the following stage, but DC coupling is even better.

Together, these noise and input capacitance considerations all but eliminate the ECC83, 6SL7GT, and other high μ, low *gm* valves from the input stage of an RIAA stage.

Strays and implementing RIAA

Our previous equalization arguments dictate that we need three individual stages to make up the complete RIAA stage, and we now know that our input valve is likely to be a high *gm* valve such as an E88CC, or better. A cascode or a μ-follower input stage are still possibilities, but we shall initially use a common cathode triode stage for simplicity. The second stage can be the same, but the third needs to be a cathode follower for reasons that will become apparent shortly. We can now draw a circuit diagram for the complete RIAA stage. See Fig. 7.23.

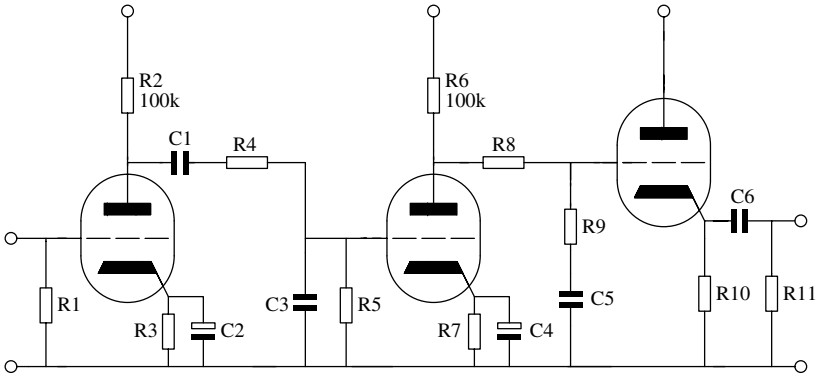


Fig. 7.23 *Basic RIAA pre-amplifier*

The 75 μs HF loss is formed by the combination of R_4, R_5 and C_3, whereas the 3180 μs, 318 μs pairing is formed by R_8, R_9 and C_5. The calculation of these components is simple, but we must remember to account for hidden components such as the output resistance of the valve, and the Miller input capacitance of the next valve in parallel with strays.

Calculation of component values for 75 μs

For the DC conditions chosen for our common cathode triode input stage, $r_a = 6\,k\Omega$, this is in parallel with the 100 kΩ anode load resistor, so $r_{out} = 5.66\,k\Omega$.

The gain of the stage is 29, and $C_{ag} = 1.4\,pF$, so the Miller capacitance will be $30 \times 1.4\,pF = 42\,pF$. In addition to this, the cathode, the heaters and the screen are at earth potential, and are in parallel with this capacitance. $C_{g-k+h+s} = 3.3\,pF$, and we ought to allow a few pF for external strays. A total input capacitance of 50 pF would be about right.

To calculate the capacitor needed for the 75μs time constant, we need to find the total Thévenin resistance that the capacitor sees in parallel. See Fig. 7.24.

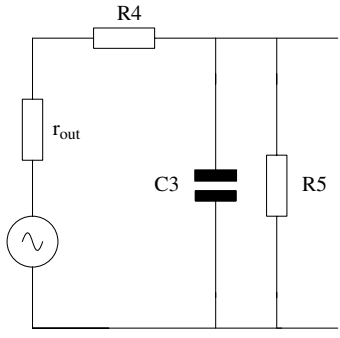


Fig. 7.24 *Determining 75 μs RIAA values*

For the moment, we will ignore C_1, but this will be accounted for later. C_3 sees the grid leak resistor R_5 in parallel with the series combination of the output resistance of the preceding valve and R_4. As is usual, we will make the grid leak as large as permissible, so $R_5 = 1\,M\Omega$.

We are now free to choose the value of R_4. We need r_{out} to be a small proportion of R_4, otherwise variations in r_a will upset the accuracy of the equalization, but too large a value of R_4 would form an unnecessarily lossy potential divider in combination with R_5. At HF, the capacitor C_3 is a short circuit, and so the additional AC load on the input valve will be R_4. 200 kΩ is a good value for R_4, and it has the bonus of being available both in 0.1% E96 series and 1% E24 series (very few E24 values are common to the E96 series). In combination with R_5, this gives an acceptable loss of 1.6 dB, whilst not being an unduly onerous load for the input stage.

The capacitor now sees $200\,k\Omega + 5.66\,k\Omega$ in parallel with 1 MΩ, which gives a total resistance of 170.58 kΩ. Dividing this value into 75 μs gives the total value of capacitance required, = 440 pF. But this network is loaded by the second stage which already has 50 pF of input capacitance from grid to ground, so the *actual* capacitance that we need is $440\,pF - 50\,pF = 390\,pF$, so a

390 pF 1% capacitor would be fine. (Cynics will think, 'How convenient that this came out as an E12 standard value.' Considerable effort was expended in nudging values so that this desirable state of affairs occurred!)

We ignored the effect of the coupling capacitor C_1, but this must have some effect on the Thévenin resistance seen by the capacitor. We could use a sufficiently large value to make its reactance negligible compared to the 200 kΩ series resistor, but a more elegant method is to move its position slightly. See Fig. 7.25.

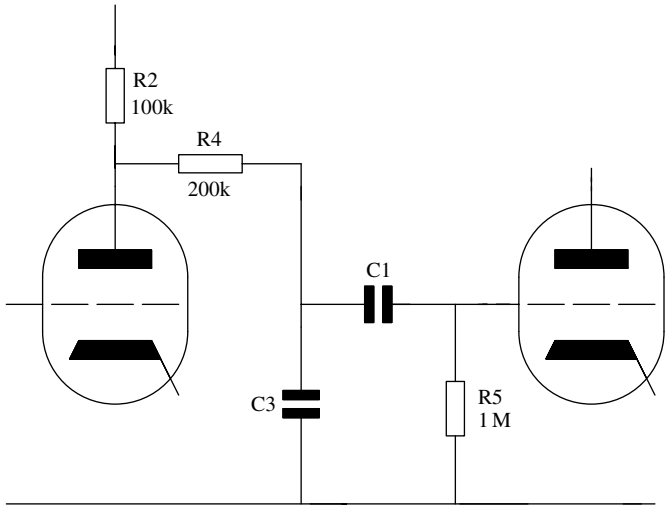


Fig. 7.25 *Moving the coupling capacitor to reduce interaction*

The capacitor now only has to be negligible compared to 1 MΩ. 75 μs corresponds to a −3 dB point of ≈2 kHz, so it is at this frequency that the values of other components are critical. At 2 kHz, a 100 nF capacitor has a reactance of ≈800 Ω, which is less than 0.1% of 1 MΩ. If we had not moved the capacitor, we would have needed a value >470 nF simply to avoid compromising RIAA accuracy.

Conversely, there is little point in using a very large coupling capacitor in an effort to reduce noise at low frequencies, since the 200 kΩ series resistance of R_4 swamps the output resistance of the input valve and nullifies its shunting effect on the grid leak of the second valve.

The following QBASIC program allows you to experiment with resistor values to achieve a convenient capacitor value and also gives the loss at 1 kHz.

```
CLS
X = 1
PRINT ''This program finds the value of capacitor''
PRINT ''required for 75us CR in the circuit of''
PRINT ''Fig. 25 and calculates the loss at 1kHz.''
PRINT
DO WHILE X > 0
PRINT ''What value is the series resistor PLUS the''
PRINT ''output resistance, in Ohms, input''
PRINT ''0 to stop'';
INPUT X
IF X = 0 THEN END
PRINT ''What is the value of the grid-leak, in''
PRINT ''Ohms'';
INPUT Y
R = (X * Y) / (X + Y)
C = .000075 / R
PRINT
D = (((C * 10 ^ 12) * 10) \ 1) / 10
PRINT ''C ='' ; D; ''pF''
A = 1 / (2 * 3.14159 * 1000 * C)
L = ((8.68589 * LOG((Y / (X + Y)) * (A / SQR R ^ 2 + A ^
 2))) * 10) \ 1) / 10
REM THE 8.68589 FACTOR ARISES BECAUSE QBASIC USES
NATURAL LOGS
PRINT ''1kHz loss ='' ; -L; '' dB ''
PRINT
LOOP
```

3180 μs, 318 μs equalization and the problem of interaction

The second stage is direct coupled to a cathode follower in order to eliminate interaction between any coupling capacitor and the 3180 μs, 318 μs pairing. 3180 μs corresponds to $f_{-3\,\mathrm{dB}} = 50$ Hz, which is far too close to our typical 1.6 Hz cut-off, so they would interact significantly.

The other reason for using a cathode follower is its low input capacitance, which causes an additional high frequency roll-off when placed in parallel with the 3180 μs, 318 μs pairing. In the 75 μs network, we were able to

incorporate the value of stray capacitance into our calculations, but in this instance it is not possible, so it is essential that any stray capacitance is so small that it can be ignored. The full equation for the input capacitance of a cathode follower is:

$$C_{input} = C_{ag} + (1 - A_v) \cdot C_{gk}$$

To a good approximation, $A_v = \mu/(\mu + 1)$, so for an E88CC ($\mu \approx 32$), $A_v = 0.97$, $C_{ag} = 1.4$ pF and $C_{gk} = 3.3$ pF. The C_{gk} term is thus entirely negligible at 0.1 pF, so the input capacitance is virtually independent of gain at $\approx$8 pF, including a healthy allowance for strays to ground.

Parallels with oscilloscope design

Tektronix oscilloscope design assumed that an E88CC/6922 cathode follower had an input capacitance of between 1 pF and 2 pF at audio frequencies, but oscilloscopes optimize bandwidth at the expense of dynamic range, so they planned their construction to minimize stray capacitance, rather than optimize screening. All electronic design can be viewed as an information problem whereby:

$$\text{Information} = \frac{\text{bandwidth}}{\text{permissible error}}$$

0.3% error is imperceptible on a CRT, but 20 MHz of bandwidth is considered paltry, whereas audio requires a very similar amount of information at a lesser bandwidth (96 dB S/N and $f_{-3\,dB} = 131$ kHz). Thus, audio design can learn much from oscilloscope and video design techniques.

3180 μs and 318 μs equalization

The equations that govern the 3180 μs, 318 μs pairing are delightfully simple, $CR = 318 \times 10^{-6}$, and the upper resistor $= 9R$, whilst the loss at 1 kHz for this network is 19.05 dB. See Fig. 7.26.

We should now check whether our worst case 8 pF shunt capacitance is sufficiently small not to cause a problem. To do this, we need to employ a slightly circular argument.

We first say that it will *not* cause any interaction. If this is true, then the frequency at which the cut-off occurs will be so high that C in the network is a short circuit. If it is a short circuit, we can replace it with a short circuit, and

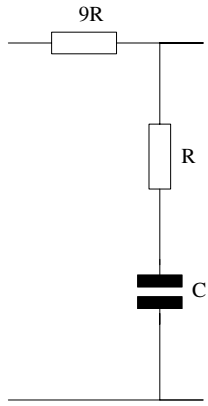


Fig. 7.26 *Implementing the 3180 μs, 318 μs RIAA pairing*

calculate the new Thévenin output resistance of the network. Since the ratio of the resistors is 9:1, the potential divider must have a loss of 10:1, and the output resistance is therefore one tenth of the upper resistor. If we assume that our upper resistor will again be 200 kΩ (neglecting r_{out} of the previous stage), the Thévenin resistance that the stray capacitance sees *at HF* is 20 kΩ, combined with 8 pF, this gives an HF cut-off of 1 MHz.

As a rule of thumb, once the ratio of two interactive time constants is ≥100: 1, the response error caused by interaction is inversely proportional to that ratio, so a ratio of 100:1 causes an error of ≈0.1 dB.

In our example, the ratio of 1 MHz to the nearest time constant of 318 μs (500.5 Hz) is 2000:1, so we can safely ignore interaction and go on to calculate accurately the values for the 3180 μs, 318 μs pairing.

If we were driving the network from a source of zero resistance, ideal values for the resistors would be 180 kΩ and 20 kΩ (perfect 9:1 ratio), since these are both members of the E24 series, and the capacitor could then be 16 nF with only 0.6% design error. Unfortunately, our source has appreciable output resistance, so we will choose 200 kΩ as the upper resistor and must accept whatever values this generates for the lower two components.

Because we used an identical stage to the input stage, output resistance is 5.66 kΩ, making a total upper resistance of 205.66 kΩ. The lower resistor will therefore be 22.85 kΩ, and the capacitor 13.92 nF. 22.85 kΩ can be made out of a 23.2k 0.1% resistor in parallel with a 1.5 M 1%. 13.92 nF can be inconveniently made out of a pair of 6.8 nF in parallel with a 330 pF. We can now draw a full diagram of the pre-amplifier stage with component values. See Fig. 7.27.

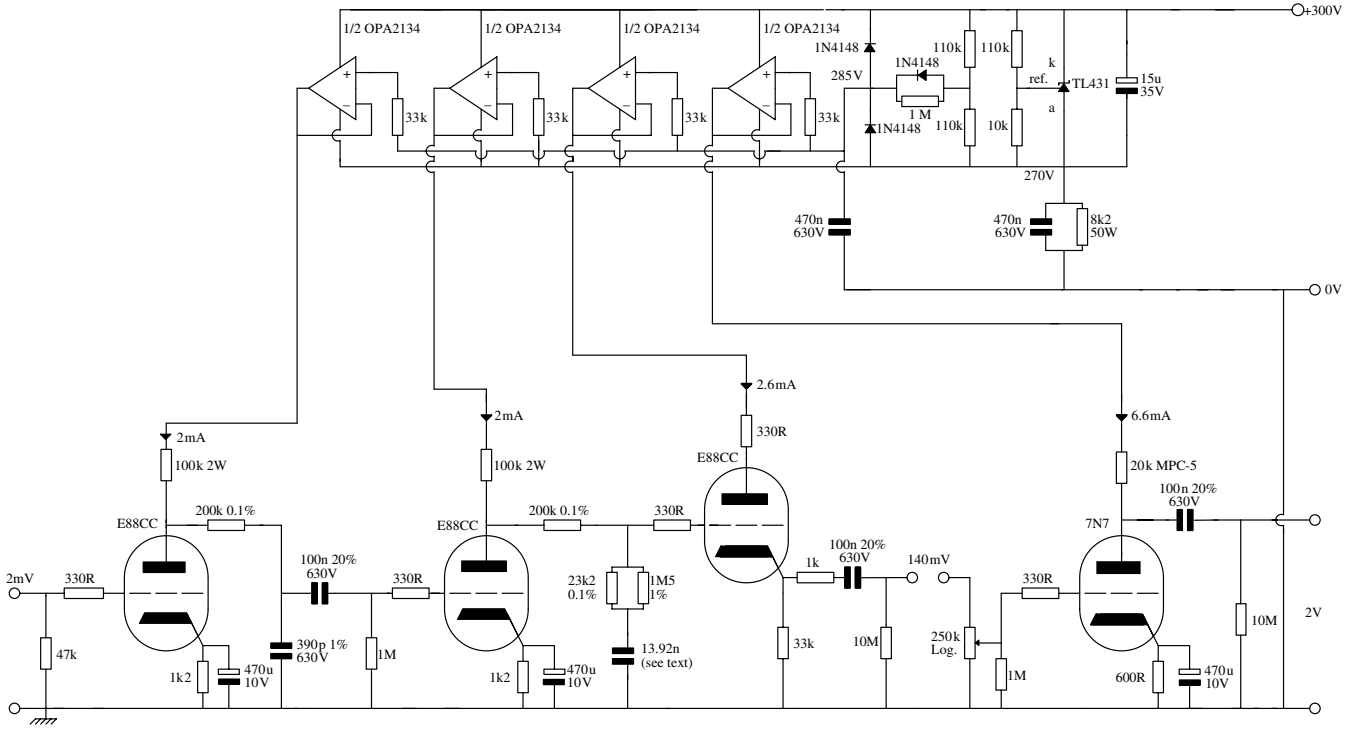


Fig. 7.27 *Practical pre-amplifier*

Awkward values and tolerances

Equalization networks and filters *invariably* generate awkward component values and much manoeuvring is required to nudge them onto the E24 series. Sadly, this is usually wasted effort, since, although 0.1% resistors are readily available, capacitors are only readily available in 1%. Therefore, for best accuracy, we measure the value of the largest capacitors on a precision component bridge (or perhaps a digital multimeter having an *accurate* capacitance range), and add an additional capacitor to achieve the required value.

For the 13.92 nF capacitor needed earlier, we might measure the 6.8 nF capacitors, and find that they were actually 6.74 nF, so we would actually need a 430 pF, rather than 330 pF. This is not a problem, but suppose we had manoeuvred a value so that exactly 10 nF was required, but when the capacitor was measured, it was found to be 10.1 nF? We can hardly file a bit off the end!

If extreme accuracy is needed, and a component bridge is available, it is better to design for values to be the nearest available value *plus* its tolerance, and then add capacitance to make up the final value. This also neatly fits our 100:1 rule for making perfect capacitors by bypassing with a smaller value.

Close tolerance components are expensive, but they are not always necessary. If we combine a close tolerance component with a looser tolerance component, then the resulting component will *still* be close tolerance, *provided* that the ratio of the values is greater than the ratio of the tolerances. Clearly, the close tolerance component must be the main component, whilst the trimming component can be looser tolerance. As an example, if we need a 22.85 kΩ resistor to a close tolerance, we could choose 23.2 kΩ 0.1%, and parallel it with 1.5 MΩ 1%. 1.5 MΩ/23.2 kΩ = 65:1, and is greater than the 10:1 ratio of the tolerances, so this combination will be fine. Similarly, for the 13.92 nF capacitor needed earlier, the ratio of the main component to the trimming component is 16:1, so even a 430 pF 10% would be fine. We could probably only buy a 1% component, so there is no need to measure it.

Just because we have adjusted component values on test to meet our exact required value *does not* mean that we now have zero tolerance components. Real components drift with time and temperature, so the values will change. What we have done is to remove the initial error, so that the practical value equals the calculated value, which places us in a better starting position for overall tolerance due to drift.

The author's prototype

The prototype of the previous pre-amplifier design was built in a case originally containing a modem, costing £3 at an electronics junk shop. A

new front panel was fitted to mask the holes in the original panel, and the case was sprayed with car paint. See Fig. 7.28.

Fig. 7.28 *The author's prototype*

Originally, the author had intended to fit the RIAA board to the turntable, but when a second Garrard 301 was acquired, this scheme was rejected, and so there are two separate signal boards within the pre-amplifier. See Fig. 7.29.

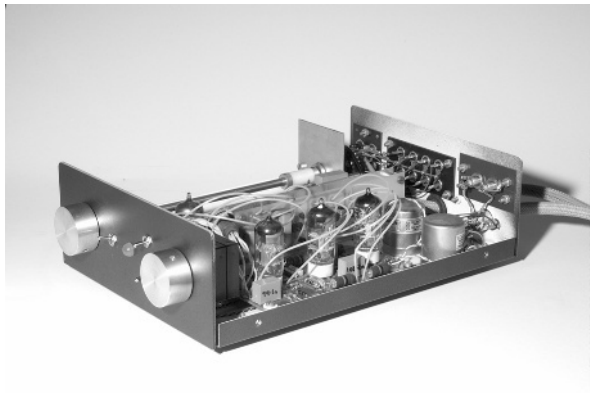


Fig. 7.29 *The author's prototype (internal view)*

The volume control was a commercially made switched attenuator (found unused in a junk shop for £1!). The Type 72 switch input selector switch mounted near the back panel has alternate contacts earthed to reduce cross-talk. Phono sockets were provided in addition to a 5-pin DIN for the moving coil disc input, which was wired to be balanced and floating (see later).

The dual colour LED in the middle of the front panel has its green LED lit by the permanently applied heater supply, and the red LED is in series with the lower leg of the HT sink resistor for the op-amps. Switching the pre-amplifier on therefore results in an orange glow similar to the colour of a valve heater, but a pure red glow would indicate LT failure.

The PCBs were ground planed, and although the pre-amplifier was initially wired with 30 strand PVC insulated extra flexible wire, this was later replaced with PTFE sleeved 0.7 mm solid silver. The change gave the best improvement for £30 that the author had heard for some time.

Balanced working and pick-up arm wiring

Balanced working is a common technique in broadcast and recording studios to protect audio signals from external electromagnetic interference, and it is particularly important for low-level signals such as microphones that occasionally have long cables (one particular television outside broadcast used a 1 km microphone cable!).

A balanced source is simply one where each terminal of the source has balanced, or equal, impedances to ground. Frequently, the only path to earth from the terminals is via stray capacitances (no DC path), and the source is then said to be *floating*. Connecting cables for balanced systems therefore have *two* identical signal wires, or *legs*, to maintain this balance, plus an overall screen. To maintain balance, the input stage of the following amplifier must have its stray impedances carefully balanced to ground, and is either based on a differential pair (cheap) or a carefully designed transformer (best, but more expensive).

When we immerse a balanced connecting cable in an electromagnetic field, an identical noise current is induced into each wire. The series resistance of the cable is the same on each leg, and the shunt capacitances and resistances to ground are also equal, so the noise current develops a voltage of identical amplitude and phase on both legs at the amplifier input. Because this is a common mode signal, it is rejected by the amplifier, whereas the wanted audio signal is differential mode, and is amplified.

A typical moving coil cartridge produces $\approx 200\,\mu V$ at 1 kHz 5 cm/s, but before RIAA equalization, the level at 50 Hz is ≈ 17 dB lower at $28\,\mu V$. Achieving our goal of inaudible hum on a signal at this level is not trivial, and we need all the help that we can get. The cartridge is inherently a balanced device, so why unbalance it?

We should immediately rewire the output cable of our pick-up arm to maintain this balance by throwing away any coaxial cable. The connecting

cable should be replaced by a twisted pair, having an overall screen for each channel. (A pair of coaxial cables for each channel would not be a good idea because the increased spacing between the inner conductors would cause slightly different noise currents in each leg, greatly reducing cancellation.)

The author's pick-up arm uses an interconnect comprising a twisted pair of 0.7 mm solid core silver threaded down a PTFE sheath, covered with a braid electrostatic screen. Both cables are then threaded down one overall braid screen which also serves to hold the cables together. All the braids are firmly bonded to the pick-up arm metalwork and mains earth at the base of the pick-up arm. The braid must not have voids, so domestic aerial cable is therefore unsuitable. Broadcast quality video cable, or multicore umbilical cable, are both ideal sources of non-voided braid. Once the plastic outer sheath has been removed, the braid easily concertinas off the inner conductors. Finally, the cable should be sleeved with nylon braid to prevent the noise that results from the cable touching another earthed metal part.

Phono plugs should *not* be used for connecting this cable to the pre-amplifier, as they are unbalanced connectors, and a 'professional' quality metal bodied 5-pin DIN plug or XLR is ideal, although the cable entry will almost certainly need to be enlarged. Alternatively, and more clumsily, a pair of 3-pin XLRs could be used, but this would require individual (double-screened) cables from the arm base, or a 'Y' split in the cable near the pre-amplifier, which is difficult to make neatly.

Within the arm tube, most pick-up arms twist all four (thin, non-screened) wires from the cartridge together, because this makes the harness easier to handle. Crosstalk between channels, and hum rejection, can be improved by twisting channels individually as they pass down the arm tube, but reverting to the four-wire twist (often required for low friction as the wires pass through the bearings to the output cable). Because this modification primarily affects longitudinal currents, it tends to be of more value to pre-amplifiers with balanced inputs, but it is still worthwhile on unbalanced ones. Martin Bastin (of Garrard modification fame) reports that he has been using this method for years.

Balanced wiring is particularly beneficial for moving coil cartridges, and even helps hum rejection when the pre-amplifier is unbalanced.

Basic pre-amplifier compromises

If we really want to achieve a significant improvement on the basic pre-amplifier, we must look closely at the fundamental design, and reconsider some of the compromises that were initially made.

- Rejection of unwanted signals was not maximized.
- Individual anode currents were set quite low in order to minimize total current consumption, allowing the pre-amplifier to be powered from an associated power amplifier. This meant that *gm* for each stage was low, and noise was not minimized.
- Metal film resistors were used as anode loads, resulting in excess noise (although most of this was shunted by r_a).
- Individual stages were kept simple, forcing poorer linearity.

A balanced pre-amplifier

The author feels that the following balanced pre-amplifier has finally reached an evolutionary dead end (although this condition does not appear to have troubled sharks over the last few million years). The ideal would be to keep the signal balanced all the way from balanced cartridge or DAC through to the push–pull output stage, but this precludes active crossovers (active crossovers are quite complex enough without making them balanced). Nevertheless, a balanced pre-amplifier makes a great deal of sense if passive loudspeakers or electrostatic headphones with a dedicated amplifier are to be used, so the design has been retained. See Fig. 7.30.

The input stage

In order to reap the full benefits of balanced working, a moving coil step-up transformer for 3 Ω cartridges was especially designed for this pre-amplifier by Sowter Transformers of Ipswich. Driven from a 3 Ω source and terminated by a pure 2.7 kΩ resistance, the first batch of type 8055 had a frequency response that was flat ±0.1 dB from 12 Hz to 100 kHz, whilst the HF phase response was pure delay ±1° to 50 kHz.

The 8055 transformer has an electrostatic screen between primary and secondary windings, and primary stray capacitances to ground have been balanced, resulting in excellent rejection of common-mode noise. As a demonstration of this rejection, the author has found that he can touch any single cartridge pin (this is the worst case), with the volume control at maximum, and not hear a sound out of the loudspeakers.

The first stage has a cascode transistor constant current sink to enhance common mode rejection. Although a 'ring of two' circuit could have been used as a sink for the first stage, each transistor would operate at a very low voltage, which not only makes the circuit more susceptible to RF overload,

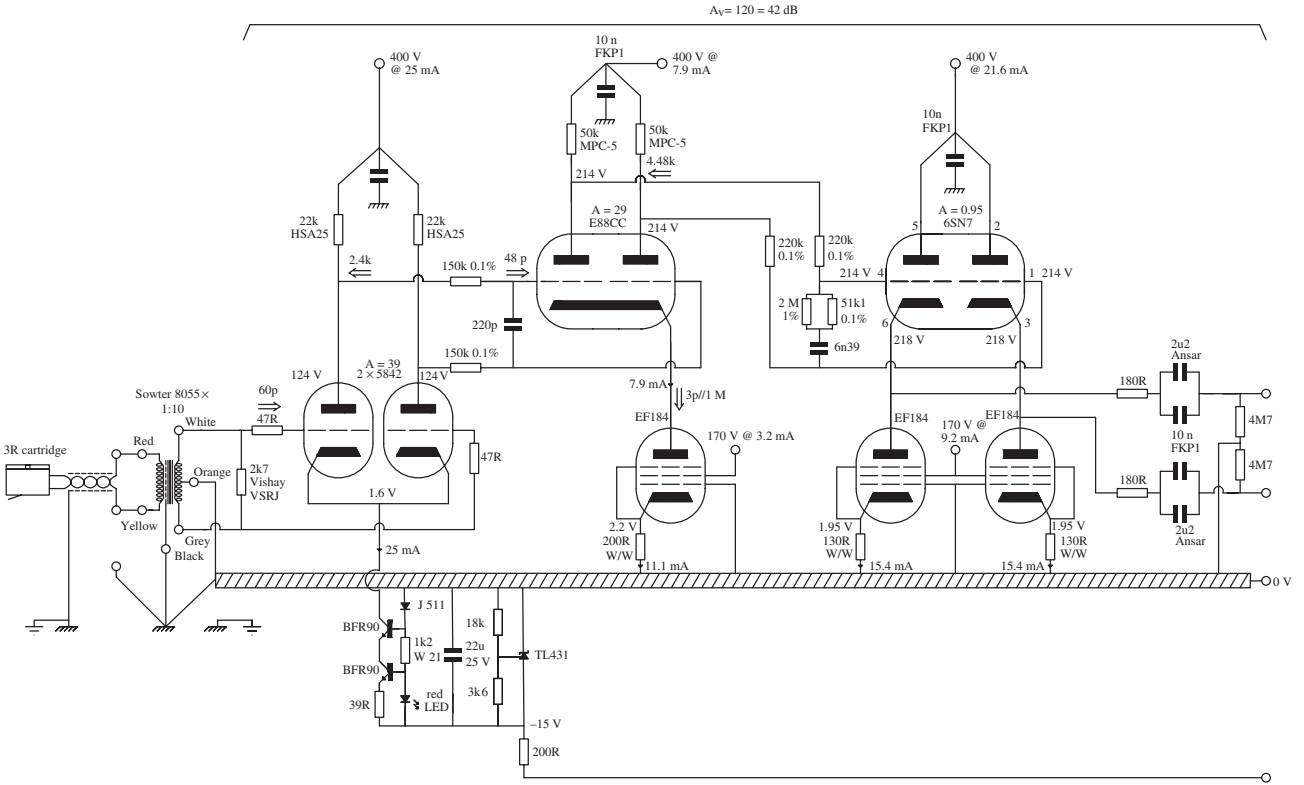


Fig. 7.30 *Fully balanced pre-amplifier*

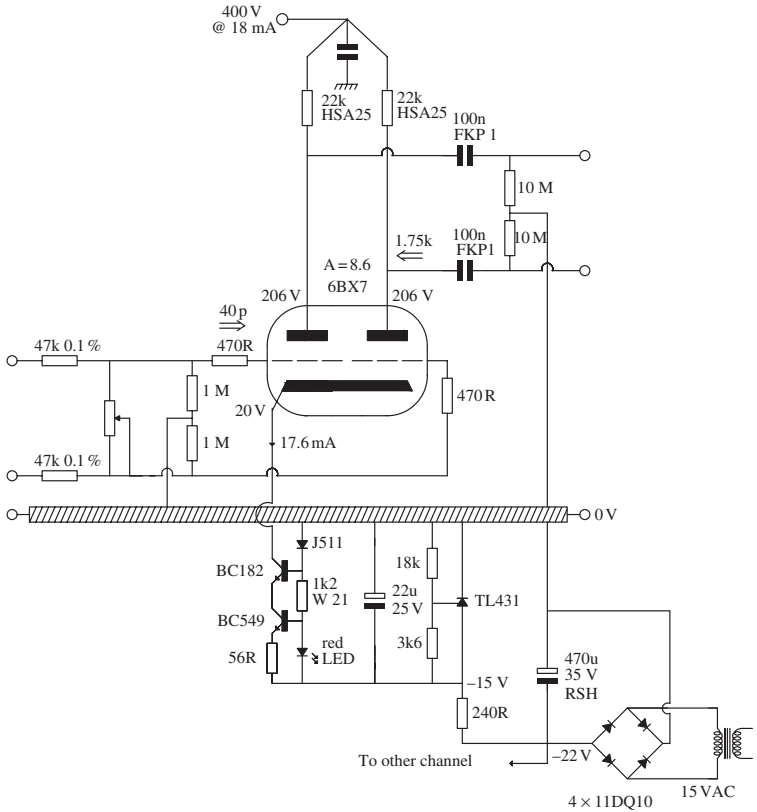


Fig. 7.30 *Continued*

but also narrows the depletion region within the transistor, increasing output capacitance. These two factors demanded the use of a subsidiary negative supply, and once this nuisance had been accepted, a superior constant current sink using RF transistors could be used, thus making a virtue out of the necessity.

The input transformer is used as the grid leak resistance for the input differential pair. The load seen by the transformer is determined purely by the loading resistor across the secondary windings, and for best sound quality this should be a bulk foil resistor. Because the differential pair is made of valves in individual envelopes, pairs can easily be found that give matched anode voltages, so a DC balance control is superfluous.

A moving magnet cartridge requires two equal value grid leaks, each of half the value required as a load by the cartridge. See Fig. 7.31.

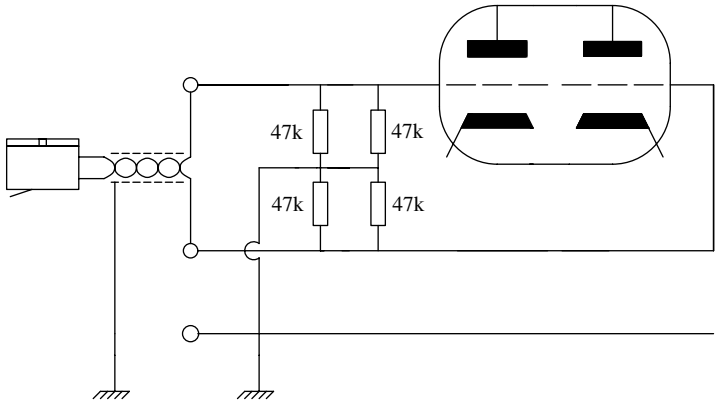


Fig. 7.31 *Arrangement of grid leaks in balanced input*

One point that may not be immediately apparent is that by connecting the cartridge in balanced fashion to both inputs of the stage, the cartridge or transformer sees the input capacitance of the valves in series, so the capacitance seen is *half* the input capacitance of one valve. See Fig. 7.32.

The second stage and 75 μs time constant

In order to direct couple the first stage to the second, the cathodes of the second stage must be at an elevated voltage, and a constant current sink seems inevitable. The excellent performance ($\approx 1\,M\Omega||3\,pF$), simplicity and cost of an EF184 constant current sink was extremely attractive, and the pentode

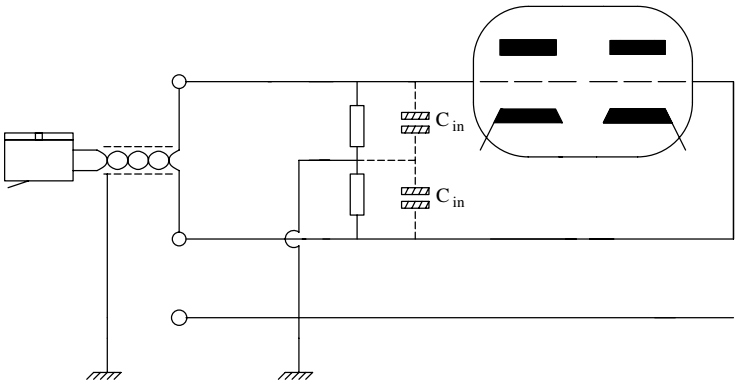


Fig. 7.32 *Reduction of Miller capacitance as seen by cartridge due to balanced connection*

noise penalty insignificant. At 20 kHz, 3 pF has a reactance of ≈2.7 MΩ, so common-mode rejection within the audio band will largely be determined by the matching of the two halves of the E88CC and stray capacitances.

Because the second stage valve is direct coupled to the first, blocking is an impossibility, so the pre-amplifier cannot convert ultrasonic spikes into sustained low frequency overload. Additionally, the second stage does not need grid leak resistors, so we avoid the 1.6 dB *excess* loss suffered by the basic pre-amplifier's 75 μs network. Every little helps.

The 75 μs time constant is achieved in a balanced fashion, and the shunt capacitor can be mounted directly on the valve base with short leads to minimize stray capacitance. If the bodies of the series resistors are as close as possible to the valve pins, this allows them to double as grid-stopper resistors. The best way to understand the equalization is to redraw the circuit as two unbalanced networks. See Fig. 7.33.

The values for *R* and C are calculated exactly as before, but we observe that if the signals are balanced, the centre tap of our added capacitors must be at AC ground, so we can disconnect it from ground with impunity, leaving two capacitors in series. These can be replaced with a single capacitor of half the value, and a noisy ground is now less able to inject noise into the audio signal. An additional advantage is that the capacitor has very little DC across it, so a lower voltage rating is permissible.

3180 μs, 318 μs pairing and associated cathode follower

Since the pairing is achieved in a balanced fashion, the value of the capacitor is halved, and it has virtually no DC across it, making it much easier to find close

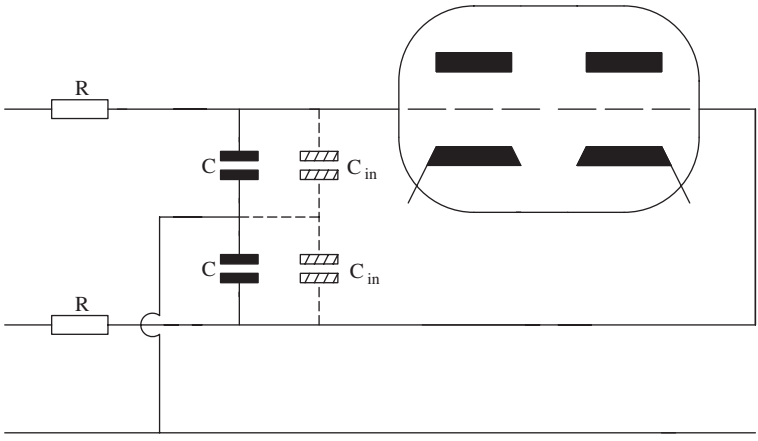


Fig. 7.33 *Understanding 75 μs implementation in balanced mode*

tolerance components. Because of the balanced 3180 μs, 318 μs pairing, twin cathode followers are necessary, resulting in a balanced output from the disc stage. The cathode followers have a rather lower V_a than is ideal for linearity, so constant current loads were substituted for resistive loads to remedy this, since an EF184 in its socket, plus current programming resistor, is cheap (unless you were the muggins who paid US $102.50 for two NOS Brimar EF184 on eBay!) 2.2 μF output coupling capacitors and high I_a in the cathode followers enable the RIAA stage to drive semiconductor equipment gracefully.

The line stage and volume control

A 6BX7 differential pair is preceded by a balanced switched attenuator (see earlier in this chapter for design details). The low μ 6BX7 minimizes Miller capacitance and r_{out}, making it useful as a line stage. The 6BX7 needs a good constant current sink because its low μ would otherwise render CMRR quite poor.

Although balanced operation cancels 2nd harmonic distortion, this theory assumes that each leg produces the same amount of distortion. Sadly, the author's recent statistical tests revealed that as a worst case, one section of a 6BX7 could produce four times the distortion of the other, destroying any hope of reliable cancellation. If you have a means of measuring distortion (perhaps a recording-quality 16-bit 48 kHz computer soundcard and appropriate software), you might want to substitute a distortion-matched pair of 6AH4 or 12B4-A rather than searching for a 6BX7 containing a pair of low distortion triodes.

Balanced interconnects and hum loops

The moving coil cartridge disc input was the obvious place to introduce a balanced interconnection because of the valuable improvement in hum and noise rejection, but balanced interconnection between the line stage and power amplifier is now becoming popular. However, we must carefully consider the 0 V signal ground and chassis connections at each end if we wish to retain maximum advantage.

'Balanced' means equal impedances and voltages to earth from either signal conductor. If the line stage has an output transformer with a centre tap, it is very tempting to connect the centre tap to earth. However, this reduces rejection of induced noise because (by definition) the output transformer cannot have an absolutely perfect centre tap, so it slightly unbalances the system. Similarly, the power amplifier input transformer (if any) must be left floating to avoid unbalancing the system.

At each end of the interconnect, the screen must be bonded directly to the chassis. Since each chassis is bonded to earth by the mains earth safety bond, this creates a hum loop. However, because the audio signal does not pass down the screen, the circulating hum current in the loop does *not* cause a problem.

Hum loops and unbalanced input stages

All unbalanced input stages are susceptible to the noise current circulating in a hum loop, but the problem can be reduced even if the loop cannot be broken.

- $V = IR$ forces the hum voltage to be proportional to the resistance of the 0 V signal conductor, so a thicker conductor reduces hum.
- If the 0 V signal ground of the amplifier is lifted away from ground, the hum current develops most of its noise voltage across the ground lift resistance rather than along the cable resistance. Lifting the ground of an amplifier assumes that it only has the opportunity to amplify the difference between the 0 V signal ground and the intended input, but flawed 0 V signal ground wiring can easily destroy this vital assumption. (Unfortunately lifting the 0 V signal ground exposes the amplifier to mains transformer HF imbalances, so practical ground lift resistances $\leq 100\,\Omega$.)
- At low frequencies, mains transformer imbalances are less of a problem, so the ground lift impedance can be increased. A 10 nF capacitor in parallel with the 100 Ω ground lift resistor is typical.

- Some designs take this concept further and treat the signal cable as a transmission line at RF. Therefore, they try to terminate the line's screen with its correct RF characteristic impedance (typically 50 Ω for RF co-axial cable, or 110 Ω/2 for twisted pair).

The perfect line stage

The best way of improving any line stage is to remove it.

It was previously argued that the ideal theoretical sensitivity for a power amplifier was $2\,V_{RMS}$, and since this is the maximum possible output of a standard CD player, the two are perfectly matched. All line stages degrade sound quality, so why not dispense with an expensive impediment, and connect signal sources directly to the power amplifier? We only need to add a volume control and input selector switch to the power amplifier to have the same facilities as before, but with much higher quality.

Modern CDs are carefully mastered so that peaks reach within 1 dB of the digital limit, but older masters might leave 3 to 4 dB of headroom. Therefore, it might be worthwhile to increase the sensitivity of the power amplifier by 3 dB from $2\,V_{RMS}$ to $1.4\,V_{RMS}$ in order to be sure that older masters are able to take advantage of the full power available.

The power amplifier is now perfectly matched to CD, but what about other sources?

- *DAB, DAT, DVD, MD, etc.* All these digital sources tend to produce $2\,V_{RMS}$, or you play them through an outboard D/A which produces $2\,V_{RMS}$, so there isn't a problem.
- *¼″ analogue tape.* If you can afford the tape costs, then you can afford a professional machine. +8 dB ref. 0 dBu $\approx 2\,V_{RMS}$, so this isn't a problem. The industry standard Studer A80 has a superb transport, but slightly less than superb audio electronics, and they are now available at very reasonable prices, making them ideal candidates for valve conversion. Studer heads need replacement once the flat across their face exceeds 3 mm, and the capstan motor shaft should not be shiny. Either replacement is expensive, so don't stand for any nonsense when bargaining.
- *Analogue cassette.* Despite the (discontinued) efforts of Nakamichi, cassette is not a Hi-Fi medium, and you wouldn't want to achieve full volume for fear of hearing all that tape noise.

- *'Hi-Fi' analogue VCR*. No, it isn't. The 'Hi-Fi' audio is recorded on an FM carrier using the rotary video heads. Unfortunately, head switching causes a momentary phase change of the FM carrier on replay, which is demodulated as a noise spike. Powerful noise reduction is employed to reduce these spikes, which occur at 50 Hz (PAL), or 60 Hz (NTSC), but it cannot remove them. Record a classical concert, replay it at a realistic level, and prove this to yourself. Once again, you won't want to achieve full volume with this source.
- *NICAM TV sound*. Although limited to a dynamic range of ≈81 dB, this is potentially quite a good source. Unfortunately, most TV tuners and VCRs have indifferent D/A converters, resulting in additional noise, so you probably won't want to achieve full volume with this source. (Some of the older tuners, such as the excellent Arcam D150, used oversampling filters that encoded the data into an SPDIF data stream, which can then be fed to an outboard D/A – see the Appendix for details of the required equalization.)
- *FM analogue radio*. Potentially quite a nice source if the broadcasters could be persuaded not to cane the Optimod (most can't). Many tuners produce quite a low audio level, and Leak Troughlines are the worst offenders. Solution: modify the decoder or tuner.
- *Vinyl*. Perhaps you don't have this, in which case, there is no problem. Most RIAA stages do *not* produce $2\,V_{RMS}$ at maximum level, and they ought.

Comparison of vinyl cartridge and CD signal levels

Although CD is specified as providing a maximum possible output of $2\,V_{RMS}$ for an undistorted sine wave, vinyl cartridges usually have their output voltage specified referred to a recorded (sinusoidal) velocity of $5\,cm\,s^{-1}$ at a frequency of 1 kHz. How can we reconcile these entirely different methods?

Vinyl is not nearly as tightly specified as CD, and maximum recorded level is totally dependent on the skill of the cutting engineer. Recorded level (and quality) may be reduced by as much as 6 dB in order to allow 40 minutes rather than 20 to be recorded on one side of a $33\frac{1}{3}$ record. RIAA equalization further complicates matters, but when equalized, vinyl typically has peaks (measured with a peak programme meter) reaching 12 dB above the nominal $5\,cm\,s^{-1}$ line-up level.

The significance of this is that we could now consider a cartridge specified to produce a nominal $2\,mV_{RMS}$ at $5\,cm\,s^{-1}$ as capable of producing musical peaks of $8\,mV_{RMS}$, which when multiplied by the 1 kHz gain of a suitable RIAA stage, could give a signal level directly comparable with that of CD.

The EC8010 RIAA stage

The main thrust of the 5842 RIAA stage in the 2nd edition of this book was to achieve a significant reduction in noise compared to the basic pre-amplifier. The EC8010 variant retains that noise improvement, but also seeks to reduce distortion significantly.

When the stylus traverses dust or a scratch, it develops an impulse having a significant DC component. At worst, when AC coupled, the overload causes blocking, but even a small DC component integrated by a cathode bypass capacitor causes a smear over time, so resistor/capacitor cathode bias will be avoided.

The input stage

The overriding requirement of the input stage is that it should produce low noise, requiring high *gm*, so the accompanying table groups valve types by *gm*.

Type	**Achievable *gm* (mA/V)**
E810F (triode connected), EC8020	≈50
3A/167M, 437A	≈42
EC8010, 5842, 417A	≈20
EC86, PC86, EC88, PC88	≈11
ECC88/6DJ8, E88CC/6922	≈8

The values given in the table are somewhat lower than manufacturers' values because they reflect usable values that can be achieved in a real design. As a very rough rule of thumb, valves designed for high *gm* typically achieve a mutual conductance of between one to one and half times their anode current. In other words, the E810F requires $I_a \approx 35\,\text{mA}$ to achieve $gm \approx 50\,\text{mA/V}$, making it expensive to use, so the choice narrowed to the $gm \approx 20$ family.

Having chosen the valve family, we must choose I_a. Since $gm \propto I_a$, we set I_a as high as is practical, so the author chose to operate the valve at $I_a \approx 15\,\text{mA}$ because this current attains most of the achievable *gm*. We next need to choose V_{gk}. Many designs have set $V_{gk} < 1\,\text{V}$, but when the author investigated the distortion spectrum of a 5842 whilst sweeping V_{gk}, he found that if $V_{gk} < 1.3\,\text{V}$, tiny changes in bias completely changed the distortion

spectrum. Once $V_{gk} > 1.5\,V$, the higher harmonics subsided and became stable, so the valve was biased by a cheap red LED (setting $V_{gk} \approx 1.7\,V$) which set $V_a = 126\,V$ for $I_a = 15\,mA$.

The value of the (wirewound) R_L can now be chosen. Theoretically, a high value of R_L increases self-noise ($v_n = \sqrt{4kTBR}$), but this is mostly attenuated by the potential divider formed by r_a and R_L, and changing the value over an extreme range only changes the final S/N ratio by $\approx 1\,dB$. The factor that determines R_L is the available HT voltage. In order to have a sufficiently large HT dropping resistor to allow adequate HT smoothing, we should keep the first stage HT $< 300\,V$, and since $126\,V$ is dropped across the valve, $174\,V$ is available for R_L, so Ohm's law determines that $R_L \leq 11.6\,k\Omega$. R_L dissipates significant power in this stage, and because a wirewound type was needed to eliminate excess noise, the nearest E6 value of $10\,k\Omega$ was chosen. (Wirewound resistors are commonly available in E6 values only.)

We now know R_L, and the current through it, so we can determine the precise HT voltage. The resistor drops $150\,V$, and $V_a = 126\,V$, so we need $276\,V$ of HT for the input stage. See Fig. 7.34.

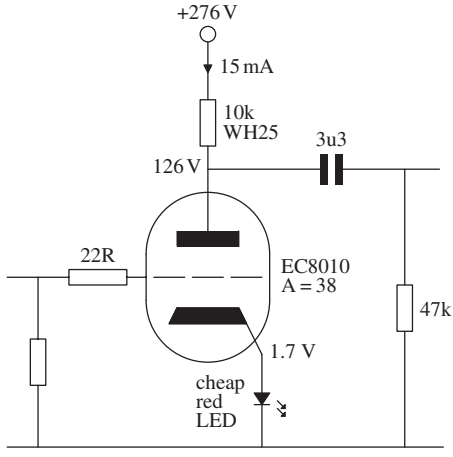


Fig. 7.34 *EC8010 input stage with LED bias*

Once the design of the input stage had been set, it could be tested for distortion. The circuit was tested at an output of +18 dBu, which lifted the distortion harmonics clear of the noise floor but was well below clipping. Twenty-six samples were tested from the EC8010, 5842, 417A family, and they were very consistent both for total THD + N and for the individual levels of their harmonics, so a typical example is shown. See Fig. 7.35.

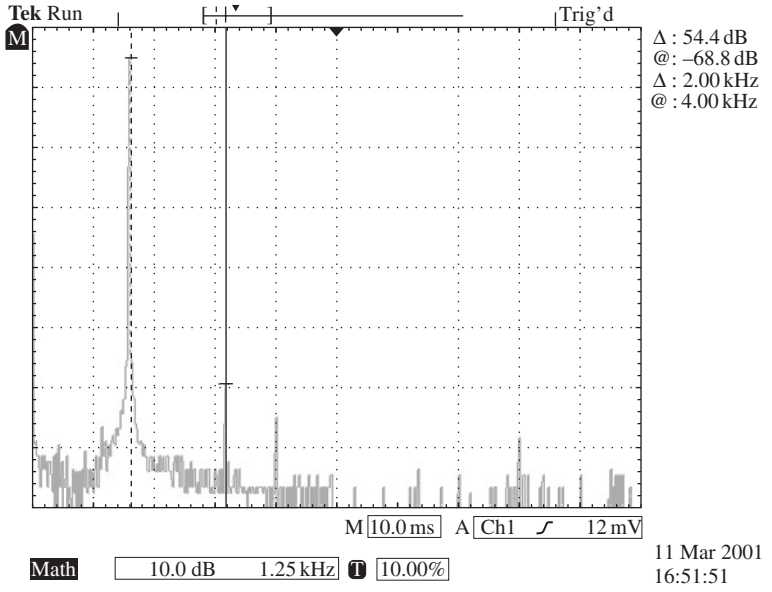


Fig. 7.35 *Typical distortion spectra of EC8010/5842/417A family at +18 dBu*

The distortion is dominated by the 2nd harmonic at −44 dB (0.65%), and the 4th harmonic is 54 dB below this at an entirely negligible −98 dB. Because distortion for a triode is proportional to level, we can predict the distortion at the proposed operating level. The nominal input sensitivity is required to be 2.5 mV_{RMS} for 5 cm/s and we convert this to dBu:

$$\text{dBu} = 20\log\left(\frac{v_{(\text{mV})}}{775}\right) = 20\log\left(\frac{2.5}{775}\right) = -50\,\text{dBu}$$

But we know that programme peaks will be 12 dB higher than this, so the peaks reach −50 dBu + 12 dB = −38 dBu. Using an EC8010, the stage had a measured gain of 32 dB, so programme peaks at the output of the stage reach −38 dBu + 32 dB = −6 dBu. We tested distortion at +18 dBu, which is 24 dB higher than −6 dBu, so the distortion at −6 dBu will be 24 dB better than that measured at +18 dBu. Thus, the distortion at −6 dBu will be −44 dB − 24 dB = −68 dB = 0.04%, which is perfectly satisfactory. If we like impressive numbers, we could instead quote the distortion at the nominal 5 cm/s level, which reduces it to 0.01%.

We should next check the input capacitance of the stage. For the EC8010, Siemens specifies $C_{ag} = 1.4\,pF$, but this will be multiplied by $(1 + A_v)$ to give a Miller capacitance of 57 pF. $C_{in} = 7\,pF$, so the total input capacitance is 64 pF. Since the author already had the stage set up on the bench, it was easy to check this value.

Adding a resistor in series with the oscillator output produces a low-pass filter in conjunction with C_{input}. Although the resistor value is not critical, so long as its exact value is precisely known. The filter $f_{-3\,dB}$ point can be found by adjusting oscillator frequency until the amplitude at the output of the test stage drops by 3 dB or its phase (relative to input) changes to 135° (180° − 45°). Using an 18 kΩ resistor, $f_{-3\,dB} = 46.9\,kHz$.

$$C_{input} = \frac{1}{2\pi f_{-3\,dB} R} = \frac{1}{2 \times \pi \times 46\,900 \times 18\,000} = 189\,pF$$

This is a long way away from the predicted value. Since we know the gain A_v of the stage, C_{ag} can be determined using the Miller equation in reverse:

$$C_{ag} = \frac{C_{input} - C_{in}}{1 + A_v} = \frac{189\,pF - 9\,pF}{1 + 38} = 4.6\,pF$$

C_{in} is the capacitance from grid to all other electrodes, as specified by the manufacturer, plus a small allowance for strays – perhaps 2 pF to 5 pF, depending on test circuit layout.

Since the manufacturer claims $C_{ag} = 1.4\,pF$, the value of 4.6 pF came as quite a surprise, but a direct measurement of C_{ag} on the component bridge gave a value of ≈4.8 pF. All bridges have trouble measuring small capacitances, and the Marconi TF2700 is no exception. Nevertheless, the manufacturer's claimed value for C_{ag} is clearly hopelessly optimistic at audio frequencies.

Optimizing the input transformer

Unfortunately, 190 pF is a large shunt capacitance for the input transformer, and initial square wave tests with the Sowter 8055 were very disappointing, but a Zobel network across the transformer secondary greatly improved matters. The required value of Zobel capacitance depends on cartridge DC resistance as shown in the table.

Cartridge R_{DC}	4 Ω	6 Ω	8 Ω	10 Ω
C_{Zobel}	1.5 nF	1 nF	910 pF	680 pF

Alternatively, the Jensen JT-346-AX transformer could be used, but we pay dearly for the (measurable) improvement in performance. The Jensen transformer was designed for 3 Ω or 5 Ω cartridges when set to 1:12 step-up ratio, so the manufacturer's data sheet gives appropriate Zobel values (assuming zero load capacitance). Experimentation revealed that 680 pF and 2k4 were optimum Zobel values for an 11 Ω source resistance and 190 pF load capacitance.

The second stage

This stage has the highest amplitude signals and therefore can be expected to produce the most distortion. As expected, when tested, gain stages with active loads such as the μ-follower and β-follower produced significantly lower distortion than a simple common cathode triode amplifier, with the further bonus of a reduced output resistance.

The μ-follower proved to be an excellent test bed for determining the irreducible distortion of a valve. The huge expansion of Internet trading means that there is now a world market for NOS valves, and almost any valve that was ever made is available somewhere. The second stage needed a valve with $\mu \approx 16$, so any likely candidate was tested – together with some unlikely ones (full details in Chapter 3).

Surprisingly, given its good reputation, the 76 did not measure particularly well. Although it produced the lowest 2nd harmonic distortion of all, its distortion was *not* proportional to level, and the higher order harmonics were at a comparatively high level. Since single-ended design relies on distortion falling with level, this valve was reluctantly eliminated.

One triode was significantly better than all others. It might be packaged as a single or a dual triode, and it might have an Octal or a Loctal base, with a 6.3 V or a 12.6 V heater, but internally the valve is the same. Perhaps unsurprisingly, this valve is the 6SN7, 12SN7, 7N7, 14N7, or 6J5. The selection was further whittled down by the discovery that the metal envelope variants produced measurably higher distortion, probably due to outgassing from the hot envelope causing increased grid ion current.

Further tests revealed that the blackened glass variants such as the CV1988 (UK military 6SN7) consistently produced the lowest distortion, but that the far cheaper Pinnacle 6J5GT was very good, and selected examples equalled the CV1988. The manufacturers claim a usefully reduced C_{ag} (3 pF compared to 3.9 pF) for the Loctals, so they can be a good alternative choice. Any of these variations on the theme would be suitable, and the final choice would probably be decided by more prosaic matters such as heater supplies, convenience of single versus dual triodes, availability of valve bases, or whether you have any.

The μ-follower and β-follower were extensively tested, and the β-follower was very good, but the μ-follower had the slight advantage that it is more flexible about HT voltage, and the author wanted to ease regulator design by operating the second and third stage at the same HT voltage.

For the Pinnacle 6J5GT μ-follower, typical distortion at +28 dBu was 0.25% or −52 dB. Using this valve, programme peaks at the output of the second stage reach +12 dBu, which is 16 dB lower, so the distortion can be expected to be −52 dB − 16 dB = −68 dB = 0.04%, which is the same as the input stage.

The μ-follower test circuit used LED bias, and although the distortion of a μ-follower is least affected by the non-constant r_{slope} of a semiconductor, this minor niggle can be removed by battery bias. A nominal 3 V lithium thionyl chloride cell actually produces an EMF of 3.6 V, which is sufficiently close to the test bias of 3.4 V (two LEDs) that it was immediately substituted. On test, no difference could be seen in the distortion spectra between battery bias and LED bias, so you could use either.

Once distortion due to the lower valve in a μ-follower has been minimized, the choice of upper valve slightly affects distortion. Various valves were tried, such as triode-strapped D3a, 6C45П and Pinnacle 6J5GT. The difference in distortion between the various types was small, but the Pinnacle 6J5GT was marginally better at 8 mA than the other valves, so it was chosen.

Having found C_{ag} of the EC8010 to be higher than expected, the Pinnacle 6J5GT was also tested. Two different measurement methods gave $C_{ag} \approx 5.4$ pF, rather than the generic value of 4 pF.

The output stage

The third stage has very similar output level requirements to the second stage, so another μ-follower was indicated. However, far less gain (and input capacitance) is required, so a 6J5 is not suitable. There are very few low

μ small triodes available – the 6BX7, 6AH4 ($\mu = 8$) and 12B4-A ($\mu = 6$) are obvious choices. These valves were designed for television field scan or series regulator use, so linearity is not guaranteed. The author had considerable qualms about the decision, but eventually decided to select from his stock of 12B4-A to find a pair of low distortion samples for the output stage.

All of these low μ valves require significant $-V_{gk}$ to set optimum operating conditions, so LED bias is impractical, and battery bias is required if bias shift is to be avoided after overload. Again, a Pinnacle 6J5GT was chosen for the upper valve.

If the 12B4-A cannot be selected for low distortion, one possible alternative is to use an NOS 37 ($\mu = 9$) for both the second stage and the output stage. Even though a test of nine samples indicated that this valve produces double the distortion of a Pinnacle 6J5GT, its distortion was far more consistent between samples than the 12B4-A, so final performance could be better than if a pair of unselected 12B4-A had to be used.

Refining valve choice by heaters

If the 12B4-A has its heaters strapped in parallel for 6.3 V ($I_h = 0.6$ A), a stereo pair requires $I_h = 1.2$ A. If a *SN7/*N7 is shared between the stereo channels for the second stage, it requires $I_h = 0.6$ A. Together with the EC8010 ($I_h = 0.28$ A), a total of 2.08 A is then required from the 6.3 V regulator. This is achievable, but a little awkward, and a pair of series 300 mA heater chains would be more convenient. Shunting the EC8010 by a 315 Ω resistance allows it to be used in a 300 mA chain, the 300 mA 6J5 is directly suitable, and the 12B4-A can be used as a 12 V 300 mA heater, so the final line-up is EC8010, Pinnacle 6J5GT, 12B4-A.

Apart from the relaxed requirements of the heater regulator, a series heater chain has many other advantages, which are detailed in Chapter 4, not least of which is reduced sensitivity to RF noise.

Achieving the RIAA equalization

We know that we will use split passive RIAA equalization, and we know the topology of individual gain stages. We must now choose impedances for the equalizers that give the best balance between distortion due to loading or grid current, and equalization errors due to stray capacitances and non-zero source resistances.

Grid current distortion and RIAA equalizer series resistances

All valves source some grid current. When a valve is fed from a non-zero source impedance, its grid current develops a voltage across that impedance. Unfortunately, this voltage (which is in series with the wanted signal) is usually distorted, so it adds distortion to the wanted signal.

Passive RIAA stages must include series resistance to form the equalizers, so this provides a mechanism for grid current to introduce additional distortion. Sadly, reducing series resistance in order to reduce grid current distortion has snags:

- At frequencies when an equalizer provides maximum attenuation, the previous stage must drive a load equal to the series resistance. Reducing a stage's load resistance steepens its loadline and increases distortion. Stages including a cathode follower, such as the proposed μ-followers, are more tolerant of loading, but caution is still needed.
- The required capacitances for the RIAA equalizers become rather large. Fortunately, 1% polypropylene capacitors are now available, but their restricted range of values means that a certain amount of juggling is necessary.

From the input stage to the second stage, a 20 kΩ series resistor would have been ideal from the point of view of grid current distortion, but this loading would have reduced the gain and increased the distortion of the EC8010 input stage. On test, 47 kΩ series resistance was a suitable compromise that minimized distortion due to the two effects. Happily, the 6J5/6J5 μ-follower second stage could comfortably drive 20 kΩ, making grid current distortion due to the third stage 12B4-A/6J5 μ-follower invisible.

3180 μs, 318 μs pairing errors due to Miller capacitance

For the basic pre-amplifier, we argued that the only logical third stage for a valve RIAA stage was a cathode follower. This was because it allowed DC coupling and had low input capacitance, which eliminated interaction and errors to the 3180 μs, 318 μs pairing. If we could tolerate interaction, and had a means of predicting and solving the problem, then this would allow a little more freedom of design choice.

If we want to achieve levels from vinyl comparable to those from CD, we *must* increase the gain of the RIAA stage. Increasing the μ of the valve at the input or second stage causes Miller capacitance problems, so the only practical way of substantially increasing gain (without increasing the number

of valves producing distortion) is to substitute a common cathode amplifier for the final cathode follower, which immediately introduces two new problems:

- The final stage must have its input AC coupled, causing interaction between the new LF roll-off introduced by the coupling capacitor and the 3180 μs time constant, producing LF response errors.
- Because the new final stage has a gain >1, Miller capacitance becomes significant, and the equalization network will be loaded by a far larger stray capacitance than before, causing HF response errors.

The 75 μs problem

Whenever possible, extended foil polystyrene capacitors are desirable for equalization networks, since this form of construction lowers self-inductance and ESR. Unfortunately, commercially available types have a voltage rating of only 63 V_{DC}, so the interstage coupling capacitor between the first and second stage has been forced to revert to its more traditional position, ensuring interaction with the 75 μs equalization.

Additionally, the grid leak resistor has been moved so that it is no longer near the grid but discharges the grid via the series resistor of the RIAA network, thus eliminating the potential divider that caused 1.6 dB loss in the basic pre-amplifier. To the author's knowledge, the first use of this cunning trick was in Arthur Loesch's transformerless RIAA MC stage.[5] See Fig. 7.36.

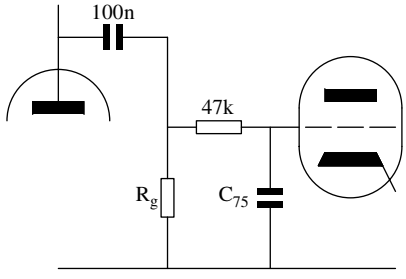


Fig. 7.36 *Modification to 75 μs implementation that eliminates unnecessary loss*

The Computer-Aided Design (CAD) solution

The various interaction problems can be solved by iterative CAD AC analysis. We start by calculating values in the normal way (assuming no

interaction), then use CAD to predict the effects of interaction on frequency response using a sweep between 2 Hz and 200 kHz. Once a problem is revealed, we adjust individual component values to seek improvement. Although this sounds laborious, it can actually be quite quick, provided that we think about how, and where, we make our adjustments.

We have five variables that must be juggled to produce the correct result, so some simplification is needed. We can start by analysing a design that does not have interaction and gently modify it, gradually introducing interactions until we reach our final design. The 3180 µs, 318 µs pairing is most affected by interaction, so we should change these components first.

3180 µs, 318 µs pairing manipulation

- The shelving loss at frequencies <20 Hz caused by adding the interstage coupling capacitor can be cured by reducing the value of the upper resistor in the potential divider.
- A mid-range shelved response (where frequencies above 1 kHz are at a constant, but different level to those below 250 Hz) can be cured by changing the lower resistor value in the potential divider. If the higher frequencies are at too high a level, this is because the potential divider has insufficient attenuation, so the lower resistor must be reduced in value, and vice versa.
- A peak in the response centred near 500 Hz can be cured by increasing the capacitor value, whereas a dip can be cured by reducing capacitor value. This result is not quite so easily deduced, but a larger capacitor would increase the time constant, lowering the frequency at which the potential divider takes effect, so that attenuation begins earlier than it should, resulting in a dip in the final response.

The last two adjustments are highly interactive, and an increase in one immediately requires a proportionate decrease (have a calculator handy) in the other to maintain the correct time constant. It is usually easiest to optimize the resistor first. The model should be tested down to 2 Hz, and the LF roll-off adjusted to emulate a simple 6 dB/octave filter, then optimized for minimum amplitude deviation from 20 Hz to 20 kHz.

75 µs/3.18 µs manipulation

Although RIAA record equalization is *specified* with only three time constants (3180 µs, 318 µs and 75 µs) this would imply a 6 dB/octave rising response at

the fragile cutting head. Allen Wright[6] points out that at the time of cutting, RIAA pre-emphasis cannot continue indefinitely and that a final time constant of ≈3.18 μs is commonly added to prevent excessive amplitude at ultrasonic frequencies from damaging the (probably Neuman) cutting head. Unfortunately, the value of this time constant varies between cutting head manufacturers, and the less common Ortofon heads use a time constant nearer to 3.5 μs. Nevertheless, it seems reasonable to accept that an electrical 3.18 μs time constant has been deliberately added at the cutting stage in addition to the inevitable mechanical losses within the cutting heads themselves. The new replay equation is therefore:

$$G_s = \frac{(1 + 318 \times 10^{-6} \times s)(1 + 3.18 \times 10^{-6} \times s)}{(1 + 3180 \times 10^{-6} \times s)(1 + 75 \times 10^{-6} \times s)}$$

Where $s = j\omega$ and $\omega = 2\pi f$.

This is an even more unpleasant equation than the original RIAA equation, necessitating a modified QBASIC program, whose detailed results are given in the Appendix. Briefly, the effect of the new equalizer is that instead of tending towards a 6 dB/octave low-pass filter, it tends towards ≈27.5 dB attenuation that is constant with frequency. Within the audio band, the new equalizer corrects a 0.64 dB loss at 20 kHz.

The justification for adding a 3.18 μs time constant to the replay equalization has little to do with amplitude response, but a great deal to do with group delay and transient response. Uncorrected, the 3.18 μs time constant changes the phase of frequencies above 5 kHz, so that they no longer arrive at the same time as lower frequencies (unequal group delay), and this distorts the transient response. We cannot compensate for the cutter, and we probably don't have the data to compensate for the cartridge response, but we can certainly compensate for the hidden 3.18 μs time constant.

An ideal RIAA stage should therefore include a final time constant of 3.18 μs, which is physically easily included by adding a resistor in series with the capacitor producing the 75 μs time constant. See Fig. 7.37.

Sadly, setting the exact value of the resistor is considerably more difficult because there are so many other HF roll-offs within the pre-amplifier, usually dominated by the output stage loading of the 3180 μs, 318 μs pairing. Usually, only the additional resistor needs adjustment, but minor adjustments of the 75 μs capacitor are likely. The model should be tested up to at least 300 kHz, and finally adjusted for optimum group delay, then checked for deviations between 20 Hz and 20 kHz. It may even be necessary to make minor changes to the 3180 μs, 318 μs pairing.

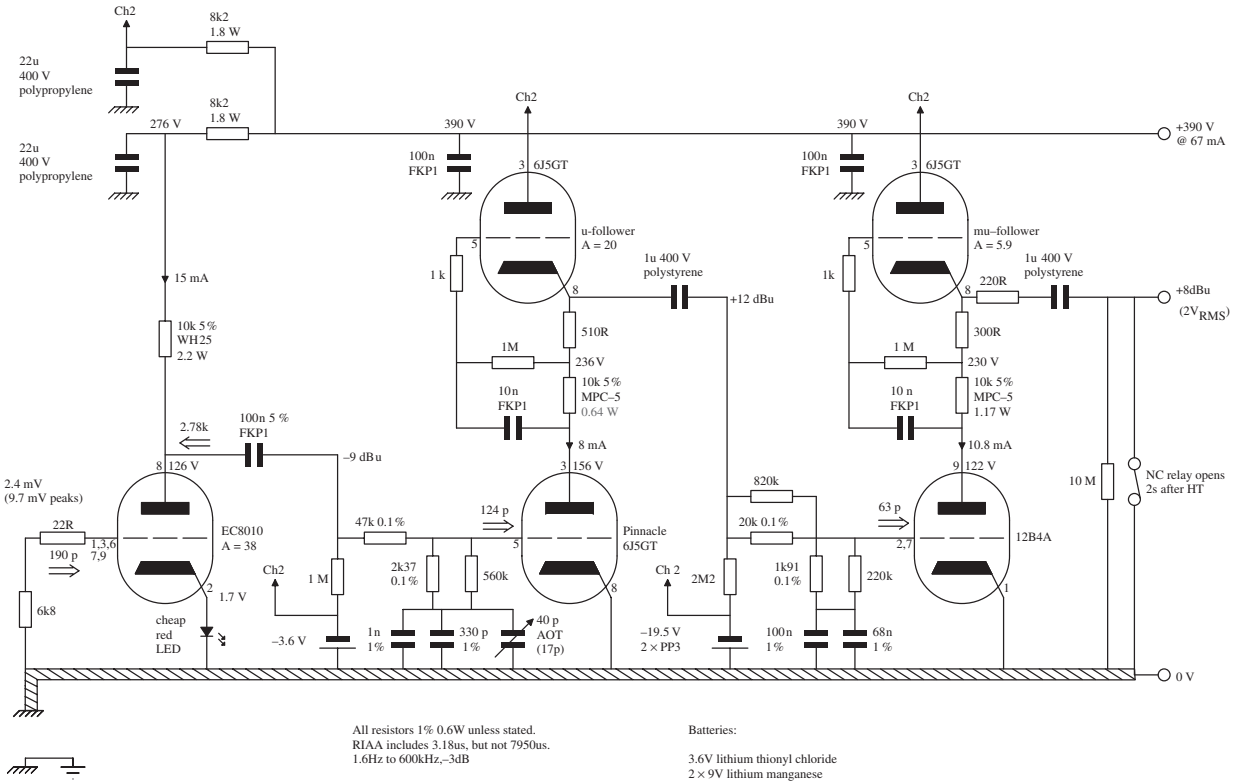


Fig. 7.37 *Final design of EC8010 μ-follower RIAA pre-amplifier*

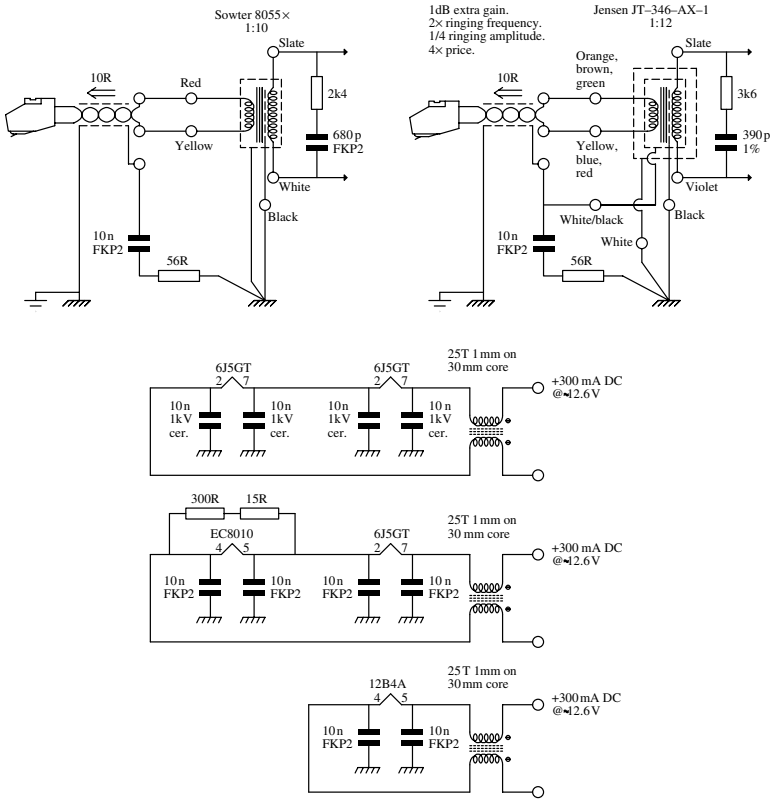


Fig. 7.37 *Continued*

Practical RIAA considerations

Setting the precise practical value of capacitance for the 75 µs, 3.18 µs pairing is awkward, so an Adjust On Test (AOT) trimmer is included. There are various alternatives for setting the AOT trimmer:

- Set the vanes almost half open (≈ 17 pF), and assume correct values for the other capacitors.
- Measure the other 75 µs, 3.18 µs capacitors on a bridge, and set the trimmer to give a predicted total capacitance of 1.35 nF, or connect them all in parallel and adjust the trimmer to give 1.35 nF.
- Measure RIAA frequency response accuracy (with a 3.18 µs capable instrument), and set the trimmer for correct response.

Although the bridge methods are indirect, they are likely to be the most accurate.

RIAA direct measurement problems

Given a well-equipped laboratory, direct measurement of RIAA equalization errors seems simple. Unfortunately, RIAA levels vary from ≈ -20 dB at 0 Hz to ≈ 25 dB at > 50 kHz making precise measurement difficult.

If a constant level is applied to the RIAA stage, the measuring amplifier must accommodate ≈ 45 dB range without any error. Conversely, setting the output level to be constant requires that the oscillator can set exact levels over a ≈ 45 dB range that can be measured precisely. Depending on the test equipment, this is either a conversion problem between the analogue and digital domains, or an analogue attenuator problem. Either way, guaranteeing attenuator error ≤ 0.02 dB and *simultaneously* a flat frequency response over a 45 dB range is not trivial, and costs money.

A popular alternative is to feed the RIAA stage via a passive RIAA pre-emphasis network and measure the combined frequency response. A theoretically ideal perfect RIAA pre-emphasis network would have an output rising indefinitely at a rate of 6 dB/octave from ≈ 5 kHz, but practical passive networks *must* have a final time constant – even if it isn't 3.18 µs.

RIAA pre-emphasis networks are quite difficult to design. Unfortunately, CAD packages can generate sufficiently large computational errors that cascading a theoretically ideal RIAA network with the inverse network whose values are to be determined iteratively can

produce design errors of ±0.05 dB. Component tolerances inevitably add further errors.

Even a perfectly designed and constructed RIAA pre-emphasis network has problems because it is sensitive to source and load impedances, which are generally considered to be constant during its design. Sadly, the carefully optimized real-world loading required by a moving magnet cartridge or a moving coil transformer disturbs the load impedance, and incorrect oscillator source resistance would compound the problem.

Summing up, keeping RIAA measurement errors below design errors is difficult.

Production tolerances and component selection

Once we have optimized component values, we can test the effects of tolerance errors. There is little point in specifying close tolerance components in one position if others with looser tolerances are able to upset performance.

The computer predicted the 20 Hz–20 kHz frequency response 10 000 times, each time with random changes in all component values within their manufacturer's tolerance. This technique is known as Monte Carlo analysis, and provided that sufficient runs are used, it predicts a likely worst case spread of frequency response. The predicted error spread for the EC8010 pre-amplifier was ±0.25 dB using the specified standard component values and without deliberate pre-selection to obtain optimum values other than setting the 75 µs, 3.18 µs trimmer capacitor to its nominal value of 17 pF.

Although RIAA errors are difficult to measure, the problem can be side-stepped by pre-selecting capacitors using a (much cheaper) component bridge, whilst a 4½ digit DVM might even allow matching of 0.1% resistors. Even without component pre-selection, the error with new valves is likely to be well within ±0.25 dB, and pre-selection could further reduce errors.

RIAA equalization errors due to valve tolerances

Even when a design deliberately sets out to minimize the effects of valve tolerances, the valves still dominate RIAA errors because passive components can now be so precise.

Unfortunately, r_{out} of the input stage is a significant proportion of the series resistance that determines the 75 µs, 3.18 µs pairing. Despite this, the computer predicted an HF shelf loss of only 0.15 dB due to *gm* of the EC8010 falling to two thirds of its nominal value.

As *gm* falls, r_a rises, which reduces gain and Miller capacitance. In this design, the valves expected to affect RIAA accuracy due to changes in Miller capacitance are the second and final valves, but as they are operated as μ-followers, changes in r_a do not affect gain ($R_L \approx \infty$), so this mechanism is not significant.

Because r_{out} for the μ-follower is such a small proportion of the resistance that determines the RIAA time constants, a tired valve in the upper stage of a μ-follower does not significantly affect RIAA accuracy.

This pre-amplifier's weakness is its sensitivity to variations of C_{ag} in the lower 6J5 of the second μ-follower. If C_{ag} rises by 50%, an HF shelving loss of 0.32 dB is predicted, whereas if it falls by 50%, an HF shelving boost of 0.34 dB is predicted. Happily, the pre-amplifier is immune to ±50% variations in C_{ag} for the 12B4-A because the 20 kΩ series resistor chosen for low distortion forces low impedances in the 3180 μs, 318 μs pairing.

Some pre-amplifiers using passive equalization with high μ valves, such as ECC83, have been found to sound audibly different with different makes of valve, giving rise to the belief that a Siemens ECC83 is better (or worse) than a Mullard, when it was actually differing r_a and C_{gk} causing clear errors in RIAA equalization.

A line stage

Although the author loathes line stages, he has to concede they are occasionally inevitable. If we must have a line stage, we should at least make it a good one.

The best justification for a line stage is that the power amplifiers have been sited near the loudspeakers in order to minimize the length of heavy loudspeaker cable, but that the sources have been sited next to the listening position. A line stage by the sources places the volume control conveniently to hand and allows short cables from the sources. However, the unity gain line stage must be able to drive the long cable to the power amplifiers.

Determination of quiescent current

Once cable is routed inconspicuously from one point to another, its required length quickly escalates. If we assume that the cable is 20 m long, typical cable capacitance of 100 pF/m would produce a capacitance of 2 nF. If we also assume transistor power amplifiers (!), we should add a further 1 nF, giving a total capacitance of 3 nF. If we want to restrict that 20 kHz loss to

0.1 dB when driving this capacitance, we must set $f_{-3\,\mathrm{dB}} = 131\,\mathrm{kHz}$, which requires a source resistance of $\approx 400\,\Omega$.

Almost any valve used as a cathode follower can achieve this small-signal output resistance, but the more significant problem is whether it can supply the required current without distortion. The reactance of a capacitor falls with frequency, so if we take the worst case, and find the reactance at 20 kHz:

$$X_c = \frac{1}{2\pi fC} = \frac{1}{2 \times \pi \times 20\,000 \times 3 \times 10^{-9}} \approx 2650\,\Omega$$

If we assume that the sensitivity of the power amplifiers has been perfectly matched to CD, we must deliver a maximum voltage of $2\,V_{RMS}$ across this reactance. By Ohm's law, this requires a current of:

$$i = \frac{V}{X_c} = \frac{2}{2650} \approx 750\,\mu A_{RMS}$$

When we consider loadlines, and valve operating conditions, we must work in peak currents and voltages, so $750\,\mu A_{RMS} \approx 1\,mA_{pk}$. Because this current is required by a capacitor, it forces a typical loadline to change from a straight line to an ellipse. See Fig. 7.38.

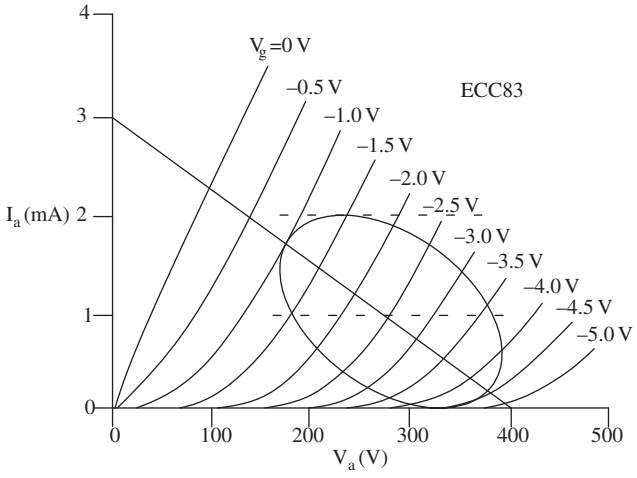


Fig. 7.38 *Elliptical loadline caused by capacitive load*

The diagram is somewhat exaggerated to aid clarity because our cathode follower will not need to swing as many volts as shown, but the effect of a capacitive load can be clearly seen. Driving the cable capacitance forces our line stage to swing vertically ±1 mA without any change in voltage (this is due to the 90° lag between current and voltage in a capacitor). As an absolute minimum, the (Class A) line stage must pass 1 mA of quiescent current so that it can swing 1 mA to the load without switching off.

Choice of valve

To achieve good linearity in the face of a heavy reactive load, a cathode follower needs plenty of distortion-reducing negative feedback, so μ needs to be as large as practicable. However, we also need the valve to have good linearity before feedback is applied, otherwise feedback will generate a spray of higher harmonics. Further, we would like *gm* to be constant with current, because we know that our elliptical loadline forces changing current. Sadly, constant *gm* in the face of changing I_a is a very tough requirement, but the valve that comes closest to these criteria is the Russian 6C45П single triode (other possibilities are triode-strapped D3a, or 6H30P). See Fig. 7.39.

Looking at the curves, we see that once $I_a > 5$ mA, the curves are nearly straight, and bunching (which reduces *gm* and μ) is almost non-existent. We

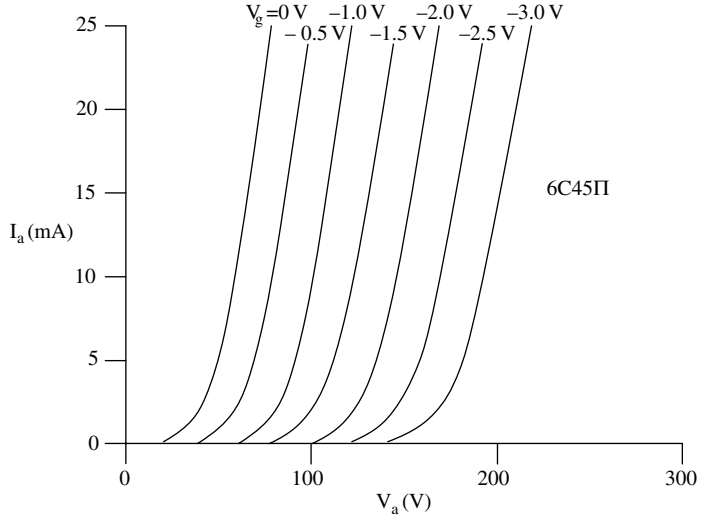


Fig. 7.39 *6C45П anode curves*

know that our quiescent current must be 1 mA clear of bunching, so we could operate the valve at $I_a = 6\,\text{mA}$, but this would be a little marginal, and 10 mA would be better.

Now that we have chosen I_a, we need to set V_a. The real limitation on V_a is that we must avoid grid current at our chosen I_a. If we choose $V_{gk} = -2.5\,\text{V}$, this is nicely clear of grid current, and sets $V_a = 170\,\text{V}$. If we use a 390 V HT (in common with the RIAA stage), we must drop $390\,\text{V} - 170\,\text{V} - 2.5\,\text{V} = 217.5\,\text{V}$, and if this passes 10 mA, a 22k load resistor would be required. Even for the 6C45П, this is a moderately steep loadline, which increases distortion before feedback, so we will use an EF184 constant current sink to force the resistive component of the loadline to be horizontal. See Fig. 7.40.

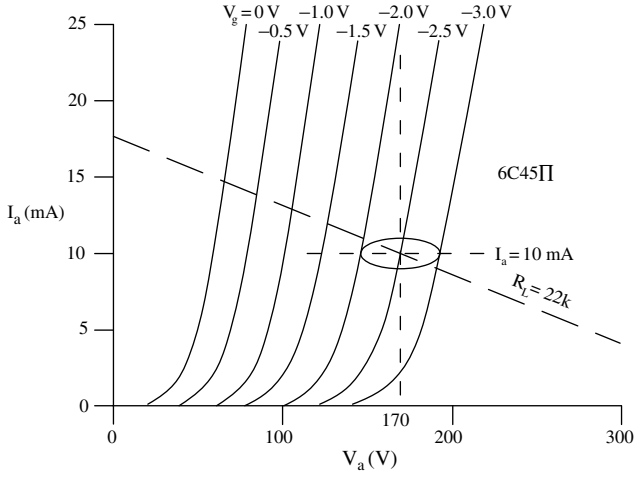


Fig. 7.40 *6C45П operating conditions in line stage*

Adding the EF184 constant current sink dramatically reduces distortion, but adds problems related to g_2. First, under these conditions, in order to sink 10 mA at its anode, the EF184 requires 4.4 mA of g_2 current, increasing the HT current for a stereo pair to almost 30 mA.

Second, if the 6C45П should fail to draw current for any reason, g_2 of the EF184 will act as an anode and will attempt to pass the full programmed current of 14.4 mA, which would destroy g_2. The g_2 problem is common to all pentode circuitry, and is commonly solved by supplying g_2 via a resistor, which limits current, so the 39 kΩ resistor protects g_2 in this circuit. See Fig. 7.41.

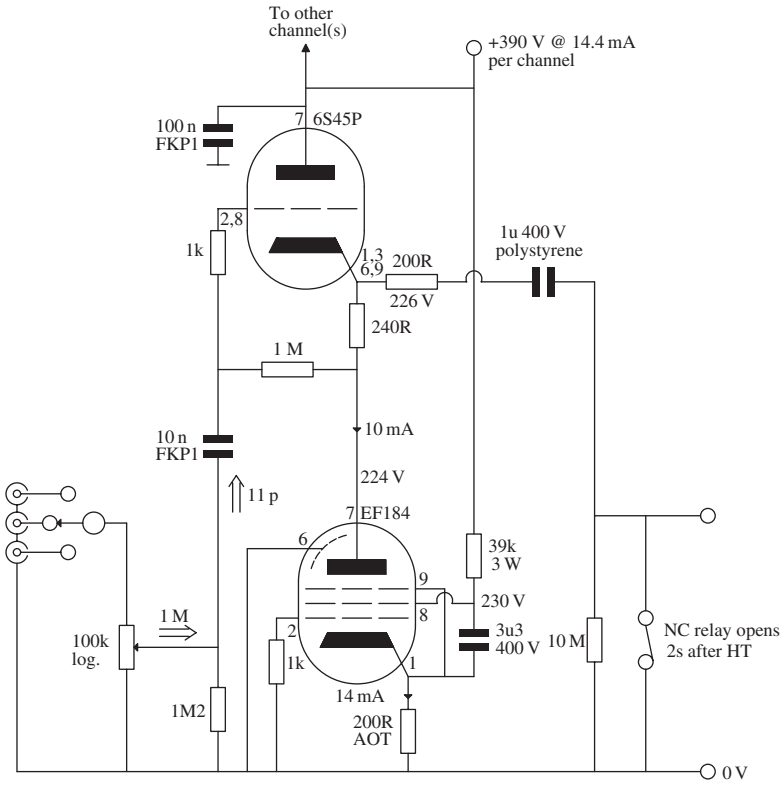


Fig. 7.41 *Low distortion unity gain line stage*

However, we can improve the performance of any pentode circuit by supplying g_2 from a low impedance supply (because I_a is far more dependent on V_{g2} than V_a), but this offers no protection against the g_2 problem. If a simple regulator such as the THINGY (see Chapter 5) was used to supply g_2, the large and expensive 3.3 μF 400 V capacitor could be replaced by a smaller collection of cheap components offering better performance, but there would be no protection against the g_2 problem. Although the EF184 is cheap, we would still prefer them to die of natural causes rather than being mugged.

If this circuit was powered in the traditional fashion using a valve rectifier and all supplies originating from one transformer, the EF184 might heat faster than the 6C45П, leaving the EF184 vulnerable to the g_2 problem. Thus, the THINGY was rejected, and the lower performance option of supplying g_2 via a resistor was adopted.

Practical considerations

The 6C45Π is self-biased by the voltage dropped across the 240 Ω cathode resistor because fixed bias from a potential divider chain would inject HT noise into the grid circuit. Because the measured input capacitance was only 11 pF, this factor does not restrict the choice of volume control, but distortion rose slightly (from 0.02% to 0.05% at +20 dBu) when the source resistance exceeded 150 kΩ, suggesting that there was grid current. A subsequent test on a modified Avo VCM163 confirmed the theory, as DC grid current was found to be constant at ≈0.1μA when V_{gk} was swept from −1 V to −3 V.

Given half a chance, the 6C45Π oscillates enthusiastically when configured as a cathode follower. The HT must be properly bypassed to ground at RF, hence the 100 nF FKP1 film/foil capacitor from the anode pin to ground. In addition, 1 kΩ grid stopper and 200 Ω cathode stopper resistors were essential to suppress oscillation at 70 MHz. If you don't have an oscilloscope that can reliably see such a high frequency at its probe tip, it might be worthwhile to raise the grid stopper to 4.7 kΩ. Similarly, because $gm \approx 16\,\text{mA/V}$ at the operating conditions, the cathode stopper resistor could be raised to 330 Ω, and still keep r_{out} below the required 400 Ω.

Passing $I_a = 10\,\text{mA}$ through the EF184 with $V_a = 221\,\text{V}$ means that it dissipates $P_a = 2.2\,\text{W}$, which is close to the 2.5 W limit, but the EF184 is cheap and plentiful, so we need not worry about a reduced lifetime. The DC conditions of the EF184 were determined in the usual way, but the value of the cathode resistor is critical to set the anode current correctly to 10 mA, so it may need to be adjusted on test. The easiest way to measure anode current is to measure the voltage across the 240 Ω 6C45Π cathode resistor using a DVM, and adjust the EF184 cathode resistor until 2.4 V is seen.

References

1 Radiolympia View, *Wireless World*, November 1949, p. 438.

2 Lipshitz, Stanley P. (1979) On RIAA Equalization Networks, *Journal of the Audio Engineering Society*, June, Vol. 27, No. 6, pp. 458–481.

3 Terman, F. E. (1955) *Electronic and Radio Engineering*, 4th ed. p. 438. McGraw-Hill.

4 Nicholson, Peter F. and Miller, Stuart (1982) *The BiFET Design Manual*, 2nd ed. Texas Instruments.

5 Few Stones Unturned. An article by Herbert Reichert featuring Arthur Loesch's RIAA MC stage. *Sound Practices* (early 1993).
6 Wright, Allen (1994) *The Tube Pre-amp Cookbook*.

Further reading

Dove, Steve (Sept. 1980 to Feb. 1982) Designing a Professional Mixing Console. *Studio Sound*.
Fletcher, Ted (Dec. 1981) Balanced or Unbalanced? *Studio Sound*.
Fletcher, Ted and Dove, Steve (Oct. 1983) Development of a Digitally-controlled Console. *Studio Sound*.
Gayford, M. (1994) *Microphone Engineering Handbook*. Focal.
Morrison, J. C. (early 1993) Siren Song: A Phono Pre-amplifier for Hedonists. *Sound Practices*.
Talbot-Smith, Michael (ed.) (1994) *Audio Engineer's Reference Book*. Focal.

Appendix

Valve data

Obtaining detailed valve data used to be extremely difficult, and the author considered himself to be very fortunate in having an almost complete set of Mullard valve data sheets. Fortunately, there are now many sources of valve data sheets, whether books or Internet sites.

Books currently in print

Mikolajczyk Piotr and Paszkowski Bohdan (1994) *Electronic Universal Vade-Mecum*. 2nd ed. Wydawnictwa Naukowo-Techniczne. 659 pages.

This beautifully presented tome is a mine of information that allows you to compare types. It doesn't have quite the detail of an original manufacturer's data sheet, but it has most of the curves, so if you only buy one data book, then this is the one to buy. Distributed by Colomor and *Hi-Fi World*.

Sylvania Receiving Tubes Technical Manual, 14th ed. 516 pages.

This book has curves for most American valves, yet the inventors of Loktal™ valves seem to be embarrassed by their offspring and consign them to the condensed data section. Reprinted by Audio Amateur Press, and available from various sources.

RCA Receiving Tube Manual Technical Series RC-19 (1959) 385 pages.

This covers many American receiving valves in good detail, and gives circuit diagrams and contemporary tips. Reprinted by Antique Electronic Supply, and available from various sources.

RCA Transmitting Tube Manual (to 4kW Plate Input). 316 pages.

All the popular directly heated audio triodes are here, but there are very few curves. Reprinted by Antique Electronic Supply, and available from various sources.

Essential Characteristics. General Electric. 473 pages.

As the title suggests, this book concentrates on breadth rather than depth, so it gives one-line information only, but it is useful, and includes nuvistors. Reprinted by Antique Electronic Supply, and available from various sources.

The Internet

The Internet has become a particularly fruitful source of data. Adobe Acrobat (pdf) files print far faster offline, so rather than launching Acrobat online and printing, save the files to hard disc and print later. (This also has the advantage that you can print new copies once the original has been obscured by loadlines.) Netscape Navigator can be configured so that it always offers the choice of launching or saving, but Microsoft Explorer requires a right mouse click every time to obtain this option.

Full manufacturers' data sheets can now be found at various sites, but the definitive site is:

http://www.tubedata.org

Frank's site is a service to the world, and the author would be lost without it.

Other very useful valve sites include:

http://members.aol.com/sbench101/
http://vacuumtube.com/

Sadly, site addresses change rapidly (in terms of book lifetimes), but many of the valve vendors provide updated links to useful sites. Keep looking.

European Pro-Electron valve codes

The Pro-Electron system of valve codes gives significant information about a valve.

	Heater type (1st letter)	Valve type (2nd and subsequent letters)		Base type (1st digit of serial number)
A	4 V	Small-signal diode	**1**	Use 2nd digit
B	180 mA	Double small-signal diode	**2**	B8B
C	200 mA	Small-signal triode	**3**	Octal
D	0.5 V to 1.5 V	Power triode	**4**	B8A
E	6.3 V	Small-signal tetrode	**5**	B9D
F	13 V	Small-signal pentode	**8**	B9A
G	5 V		**9**	B7G
H	150 mA	Hexode or heptode		
K	2 V	Heptode or octode		
L		Power tetrode or pentode		
M		Fluorescent indicator		
N		Thyratron		
P	300 mA			
Q		Nonode		
X		Gas-filled rectifier		
Y		Half-wave rectifier		
Z		Full-wave rectifier		

Example

ECC88 has a 6.3 V heater, with two small-signal triodes on a B9A base. For European manufacturers such as Mullard or Philips, reversing the order of digits and letters after the heater letter signified a Special Quality type, such as E88CC, or E180F.

It would be nice to say that the coding system was rigidly observed, but it's rather like French verbs – there are some irregular types. Marconi Osram and STC completely ignored the system and used proprietary codes that were identifiably weird, but the recently marketed 6.3 V heater, dual triode ECC99 valve actually uses a B9A base rather than the B7G implied by its code.

Ediswan/Mazda applied their proprietary code in a startlingly Humpty Dumpty fashion ('When *I* use a word, it means just what I choose it to mean – neither more nor less'), sometimes they conformed to their code, sometimes they didn't.

Heater type 1st number		Valve type 1st and subsequent letters	
6	6.3 V	**C**	Frequency changer
10	100 mA	**D**	Small-signal diode
20	200 mA	**F**	Small-signal pentode
30	300 mA	**K**	Thyratron
		L	Triode
		M	Fluorescent indicator
		P	Pentode or beam tetrode
		U	Half-wave rectifier
		UU	Full-wave rectifier

Thus, the Mazda 6F13 is a small-signal pentode with a 6.3 V heater, and a 6/30L2 is a small-signal triode having a 6.3 V 300 mA heater, but you would need the data sheet to discover that it is actually a *double* triode.

American consumer valve codes

Like the Pro-Electron system, the American RETMA number/letter system sought to introduce some logic to consumer valve codes. Unfortunately, this utopian ideal was soon overturned, and a non-systematic four digit industrial code was used in tandem. Nevertheless, the number/letter system offers some clues as to a valve's internals.

American consumer valve codes:

1st digit	Letters	2nd digit
Approximate heater voltage. (However, 7 or 14 means Loctal base.)	Purely random – relates to individual valve design.	Number of electrodes (including heater and metal envelope).

Example

6SN7 has a 6.3 V heater and seven electrodes (two individual triodes plus one heater adds up to seven). The 12SN7 is functionally identical, but

requires a 12.6 V heater because the heaters are wired in series rather than parallel. There is no other difference.

In addition, the following American Octal suffixes were also used.

American Octal valve code suffixes:

Suffix	Meaning	Comment
None	Metal envelope	Introduced in 1935. Although the envelope provides a useful screen, the envelope outgasses at high temperatures and the consequent grid gas current worsens noise and distortion performance.
G	Glass	Early valves tended to use the ST14 (Shouldered Tube) envelope that looks like a soft drink bottle.
GT	Glass, Tubular	Later glass envelopes were of a shorter, tubular construction.
GT/G	Interchangeable	Usable with equipment specified for either G or GT.

The final suffix signified the development of a basic type, and they are reverse compatible with the older type. Thus, for the 6SN7:

	6SN7GT	6SN7GTA	6SN7GTB
$P_{a(max.)}$ per triode	3.5 W*	5 W	
$P_{a(max.)}$ total	5 W	7.5 W	
$V_{a(max.)}$	300	450	
Controlled warm-up time	–	–	Yes

*Depends on date of manufacture and manufacturer. Early valves tend to be 2.5 W, later ones 3.5 W.

The increased voltage rating was usually achieved by additional perforations in the supporting micas to reduce leakage paths between anode and grid. The increased power rating was usually achieved by changing the anode from single to double sutures which increased the radiating area and allowed better cooling.

For many valves, the 'B' suffix indicates controlled warm-up time. Ideally, series heater chains should be driven by constant current sources, but in

practice, they were more likely to be connected directly across the mains supply (a constant voltage source). The heater chain is a potential divider that sets individual filament voltages correctly in normal operation, but as the filaments warm, their resistance changes, and unless all the valves change their resistance in concert, some filaments will be overstressed, eventually leading to failure.

Since most valves were intended for use in radios, capacitances and screening were important. As an alternative to the metal envelope, early glass envelope valves have a pressed metal ring (often connected to pin 1) encircling their phenolic base that acts as a guard to reduce capacitances to other components.

E88CC vs ECC88 heater current

The E88CC/6922 is a special quality version of the ECC88/6DJ8, and all electronic characteristics are identical, except that the heater is specified as 365 mA. *However*, only Mullard/Philips/Amperex seem to have adhered to this specification, and most manufacturers used 300 mA heaters. Brimar specify 365 mA, but the author tested ten Brimar ECC88/6DJ8, and found that they were all 300 mA.

Pin connections

Unlike ICs, valves are numbered *viewed from underneath* counting clockwise.

Thermionic emission

The Richardson/Dushmann equation for emitted cathode current (A) per unit area (m^2) is:

$$I = AT^2 \exp^{-\frac{e\Phi}{kT}}$$

where: $A = \left(\frac{4\pi m_e e k^2}{h^3}\right) = 1.204 \times 10^6\ A/m^2/K^2$, see note

Φ = work function of the cathode surface (≈ 4.55 for tungsten)
k = Boltzmann's constant $\approx 1.381 \times 10^{-23}$ J/K
T = absolute temperature = °C + 273.16
exp = base of natural logarithms ≈ 2.718
m_e = electron rest mass $\approx 9.109 \times 10^{-31}$ kg
e = electronic charge $\approx 1.602 \times 10^{-19}$ C
h = Planck's constant $\approx 4.136 \times 10^{-15}$ eVs

Note that although the theoretical value for $A = 1.204 \times 10^6$, the experimental value is about half this value.

Square wave sag and low frequency $f_{-3\,\mathrm{dB}}$

A square wave with LF sag is a decaying exponential, whose instantaneous voltage at any time 't' may be found using:

$$v = V_0 \exp^{-\frac{t}{\tau}}$$

Rearranging, and solving for τ:

$$\tau = \frac{-t}{\ln\left(\frac{v}{V_0}\right)}$$

't' is the time allowed for the decay across the bar top, but for a square wave with equal positive and negative durations, it is half of the periodic time T:

$$T = 2t$$

But T is the reciprocal of frequency:

$$f = \frac{1}{T}$$

So:

$$t = \frac{1}{2f}$$

Substituting:

$$\tau = \frac{-1}{2f\ \ln\left(\frac{v}{V_0}\right)}$$

From the frequency domain, a CR filter has a -3 dB cut-off frequency:

$$f_{-3\,\mathrm{dB}} = \frac{1}{2\pi CR}$$

But $CR = \tau$, and $\tau = L/R$, so a universal equation, valid for both CR and LR is:

$$f_{-3\,\mathrm{dB}} = \frac{1}{2\pi\tau}$$

Rearranging:

$$\tau = \frac{1}{2\pi f_{-3\,\mathrm{dB}}}$$

We now have two formulae for τ, which can be equated:

$$\frac{1}{2\pi f_{-3\,\mathrm{dB}}} = \frac{-1}{2f\,\ln\left(\frac{v}{V_0}\right)}$$

Solving for the ratio $f/f_{-3\,\mathrm{dB}}$:

$$\frac{f}{f_{-3\,\mathrm{dB}}} = \frac{-\pi}{\ln\left(\frac{v}{V_0}\right)}$$

Sag is the percentage of peak-to-peak level by which the horizontal bar has sagged in level. 10% sag is easily measured on an oscilloscope, and means that the level has fallen from 100% to 90% so:

$$10\%\ \mathrm{sag} \equiv \frac{v}{V_0} = 0.9$$

Applying 10% sag to the $f/f_{-3\,\mathrm{dB}}$ formula:

$$\frac{f}{f_{-3\,\mathrm{dB}}} \approx 30$$

So 10% sag means that the applied square wave frequency is 30 times higher than $f_{-3\,dB}$.

Sag observed using a square wave of frequency '*f*'	$f/f_{-3\ dB}$
10%	30
5%	60
1%	300

Playing 78s

You have just inherited a collection of 78s that appear to be in superb condition, some of them are original recordings of legendary performers, and you are desperate to play them.

There are four main problems.

Correct speed

Although colloquially known as '78s', referring to their speed, very early 78s were recorded on rather crude lathes and the actual recorded speed was somewhat variable. The first requirement for replay is therefore a turntable that will not only rotate at 78 rpm, but also has varispeed, so the Garrard 301/401 and Thorens TD124 are obvious contenders. The BBC modified the Technics SP10 direct drive turntable to give varispeed, added a pick-up arm/cartridge plus elaborate electronics, and called the whole confection an RP2/10 (ReProducer 2, version 10). The Technics SL1200 and EMT 938 turntables are also capable of playing 78s.

Groove size

The 78 has a coarse groove, and was traditionally played with a crude steel 'needle'. The LP stylus of a modern cartridge is far too small, and a dedicated large diameter stylus is required. Traditionally, only broadcast cartridges like the Shure SC45 and Ortofon OM Pro were offered with 78 styli by their manufacturers but the situation has now changed. Vinyl is now a niche market, and the cartridge manufacturers are having to respond to

the demands of that niche, with the result that almost all manufacturers now offer a mono cartridge, either for mono vinyl, or with a larger tip for 78s.

Grado goes one step further and offers a mono 78 cartridge with four different size tips to give the best compromise between worn grooves and surface damage. Alternatively, a specialist retipping concern may be prepared to fit a 78 tip to your cartridge.

Because different 78s are likely to need a different size tip, moving magnet cartridges are more suitable for playing 78s (Audio Technica OC9 notwithstanding) because removable styli are the norm. Even so, the Lyra Helikon Mono *moving coil* cartridge, could presumably be fitted with a 78 tip, and if you had an arm with a removable headshell, different cartridges could have different tips. You would need an awful lot of priceless 78s to justify the expense. Probably, only a national library that wanted to transcribe their recordings could afford the expense.

Incidentally, the 78 is the only recording medium that is provably robust over decades. Magnetic tape went through a sticky patch in the 70s with unstable binders. 'Perfect sound for ever' turned out not to be true if the CD had been pressed at Blackburn. The plant had been converted from pressing Laservision video discs to CDs, and used silver rather than aluminium as the reflective coating, which wasn't a problem, but they then had a problem with imperfect lacquer sealing at the periphery. Even that wouldn't have been too bad, except that the paper liners had an unusually high sulphur content, which promptly reacted with the silver to produce yellow silver sulphide, which isn't so reflective, and causes data errors. 'Laser rot' was nonsense. Any recording format based on videotape (CDs were originally mastered on U-matic) becomes unplayable not because the medium deteriorates, but because working machines are unavailable. 2" quadruplex videotape is only 50 years old, and was in common use for almost 30 years, yet there are now very few of even the final generation machines in working order.

As a species, we are becoming increasingly sloppy about recording our heritage. The Rosetta stone might have been difficult to read, but at least that was largely a linguistic problem, rather than deterioration of the medium.

78s *must* be played with an appropriate stylus. Because the groove of a 78 is so much larger, an expensive microgroove stylus would founder on the bottom of the groove, producing unnecessary noise, and wearing itself on the abrasive shellac as it does so.

Pick-up arm mechanics

Playing a 78 drives a lot of vibration into the pick-up arm, and loose bearings cause rattles and mistracking. At a more subtle level, a stylus traversing an

imperfection, or speck of dust, produces a mechanical impulse and excites arm resonances which greatly magnify the subjective nuisance. Paradoxically, the inferior medium needs a good arm to replay it adequately – a modern arm such as the Rega RB250 arm and its derivatives seems a minimum requirement.

Equalization

Analogue disc

It took some time before the manufacturers of 78s and LPs standardized their equalization. The following table gives the electrical time constants used by major organizations, and therefore an indication of the likely equalization required.[1]

(All time constants are specified according to Lipshitz' notation.[2])

	Time constants (μs)	t_3	t_4	t_5
78	'Standard'	–	636	–
	Decca 'ffrr'/European	–	636	25
	AES	–	400	63.6
	Pre-1954 DG	–	450	50
	BBC	–	450	25
	International	–	**450**	**50**
LP	Pre-1954 DG	1590	450	50
	Pre-1954 Decca	1590	318	50
	Columbia/EMI	1590	318	100
	European	2230	318	50
	NAB	3180	318	100
	RCA New Orthophonic	3180	318	75
	RIAA	**3180**	**318**	**75**

Errors in t_4 and t_5 cause peaks and troughs in the critical mid-band, and despite popular belief, cannot possibly be corrected using tone controls.

If only the later 78s (when an international standard had been fixed) are to be played in addition to modern LPs, then t_4 need only be switchable between 318 μs and 450 μs, and t_5 between 75 μs and 50 μs.

Langford-Smith[3] stated that the Decca (London) LP recording characteristic had the following response (Jan. 1951):

Frequency	30 Hz	50 Hz	100 Hz	300 Hz	1 kHz	10 kHz	15 kHz
Level	−17.5 dB	−14 dB	−9 dB	−3 dB	0 dB	+14 dB	+16 dB

Lipshitz time constants that produce a response passing within ±0.1 dB of these points are:

$$\tau_3 = 9.6\,\text{ms}$$
$$\tau_4 = 735\,\mu\text{s}$$
$$\tau_5 = 110\,\mu\text{s}$$
$$\tau_6 = 10.2\,\mu\text{s}$$

Modern analogue 'microgroove'

	τ_2	τ_3	τ_4	τ_5
RIAA	–	3180 μs	318 μs	75 μs
IEC	7950 μs	3180 μs	318 μs	75 μs

Note that the 7950 μs IEC time constant is *replay only*, and is a 20 Hz high-pass filter intended to remove rumble produced by turntables. See Chapter 7 for details on implementation.

The RIAA record equalization implies a 6 dB/octave rising response with frequency. If a 3.18 μs time constant is added to the head power amplifier to protect the fragile cutting head from ultrasonic energy, the Lipshitz replay equation becomes:

$$G_s + \frac{(1 + \tau_4 \cdot s) \cdot (1 + \tau_6 \cdot s)}{(1 + \tau_3 \cdot s) \cdot (1 + \tau_5 \cdot s)}$$

where:

$$\omega = 2\pi f$$
$$s = j\omega$$
$$\tau_3 = 3180\,\mu s$$
$$\tau_4 = 318\,\mu s$$
$$\tau_5 = 75\,\mu s$$
$$\tau_6 = 3.18\,\mu s$$

Thus:

$$G_s = \frac{(1 + 318 \times 10^{-6} \times s)(1 + 3.18 \times 10^{-6} \times s)}{(1 + 3.18 \times 10^{-3} \times s)(1 + 75 \times 10^{-6} \times s)}$$

Alternatively, the following equation[4] (modified to include 3.18μs) can be used:

$$G_\omega = 10\log\left[1 + \frac{1}{(\omega\tau_4)^2}\right] - 10\log\left[1 + (\omega\tau_5)^2\right] - 10\log\left[1 + \frac{1}{(\omega\tau_3)^2}\right] + 10\log\left[1 + (\omega\tau_6)^2\right]$$

The first (fundamental) equation requires considerable manipulation, but allows phase to be found, whereas the second is far easier to compute if only gain is required. The two equations produce precisely the same gain results.

Note that including the 3.18 μs time constant changes the extreme HF response from a 6 dB/octave low-pass filter to a final attenuation of 27.5434 dB that is constant with frequency.

CD

50 μs, 15 μs

This equalization is only very rarely used on CDs, and is accompanied by a sub-code flag to enable the player to apply the equalization. Many over-sampling filters can implement this equalization digitally with extreme precision, but this negates the noise advantage of pre-emphasis over the DAC, so precisely implemented analogue equalization is better.

Frequency (Hz)	Gain (dB. ref 1 kHz)	Phase (degrees°)
0	+19.909	0
10	+19.741	−10.4
20	+19.272	−20
50.05	+16.940	−40.6
70	+15.281	−48.3
100	+13.087	−54.7
200	+8.218	−59.3
500.5	+2.642	−52
700	+1.233	−48.9
1000	0	−47.8
2000	−2.583	−53.6
2122	−2.860	−54.5
5000	−8.168	−66.4
7000	−10.733	−68.9
10 000	−13.566	−69.3
20 000	−18.979	−63.5
50 050	−24.541	−43.1
70 000	−25.754	−34.2
100 000	−26.574	−25.6
200 000	−27.280	−13.6
500 000	−27.500	−5.5

NICAM 728

Some countries broadcast terrestrial analogue TV with digital audio encoded as NICAM (Near Instantaneous Companding And Modulation). The 728 refers to the data rate –728 kb/s compared to a typical channel data rate of 128 kb/s (and falling) for DAB. Beware of making a direct comparison of data rates – NICAM is a much earlier digital compression system that does not take as much advantage of pyschoacoustic masking.

NICAM 728 uses a CCITT analogue pre-emphasis characteristic known as J17, whose response curve can be calculated from the following formula:

$$\text{Loss}_{(\text{pre-emphasis})} = 10\log_{10}\frac{75 + \left(\frac{\omega}{3000}\right)^2}{1 + \left(\frac{\omega}{3000}\right)^2}\,(\text{dB})$$

This is not a useful equation for determining component values for de-emphasis (equalization), so the following equations will be found to be far more useful:

$$C = \frac{1}{3000x}$$
$$x = y(\sqrt{75} - 1)$$
$$y = \frac{x}{\sqrt{75} - 1}$$

where x, y and C are the components in the diagram. See Fig. A.1.

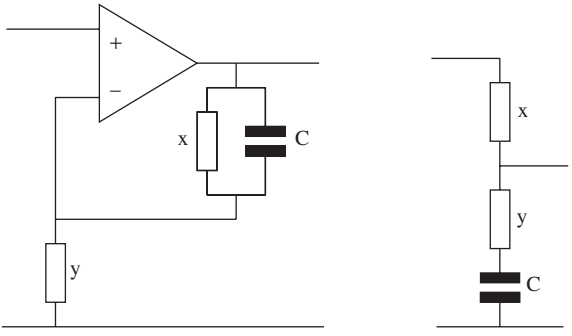


Fig. A.1 *Achieving J17 equalization*

Standard resistor values of 36 and 4.7 for x and y result in a calculation ratio error of <0.01%.

Standard component values

The following series of components covers one decade; other values are obtained by multiplying or dividing by factors of ten.

E6

1 1.5 2.2 3.3 4.7 6.8

E12

1	1.2	1.5	1.8	2.2	2.7	3.3	3.9	4.7	5.6	6.8	8.2

E24

1	1.1	1.2	1.3	1.5	1.6	1.8	2	2.2	2.4	2.7	3
3.3	3.6	3.9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	9.1

E96

1	1.02	1.05	1.07	**1.1**	1.13	1.15	1.18	1.21	1.24	1.27	**1.3**
1.33	1.37	1.4	1.43	1.47	**1.5**	1.54	1.58	1.62	1.65	1.69	1.74
1.78	1.82	1.87	1.91	1.96	**2**	2.05	2.1	2.15	2.21	2.26	2.32
2.37	2.43	2.49	2.55	2.61	2.67	2.74	2.8	2.87	2.94	3.01	3.09
3.16	3.24	3.32	3.4	3.48	3.57	3.65	3.74	3.83	3.92	4.02	4.12
4.22	4.32	4.42	4.53	4.64	4.75	4.87	4.99	5.11	5.23	5.36	5.49
5.62	5.76	5.9	6.04	6.19	6.34	6.49	6.65	6.81	6.98	7.15	7.32
7.5	7.68	7.87	8.06	8.25	8.45	8.66	8.87	9.09	9.31	9.53	9.76

Values in bold are common to E24 series.

Resistor colour code

Most resistors are marked with their value in the form of a colour code consisting of four or six concentric bands of paint on the body of the component which are read from left to right. See Fig. A.2.

Four band resistors

The first two bands denote the two significant digits of the value.

The third band is the *multiplier*, whose value is 10^x, where x is the value of the band. Gold used as a multiplier means $10^{-1} = 0.1$, and silver means $10^{-2} = 0.01$.

The fourth band is the tolerance, which will commonly be 1% (brown) or 2% (red). On older equipment, you will see gold (5%) and silver (10%); the use of these colours as tolerances dates from the days when 5% was considered to be close tolerance! If there is no fourth band, the tolerance is 20%.

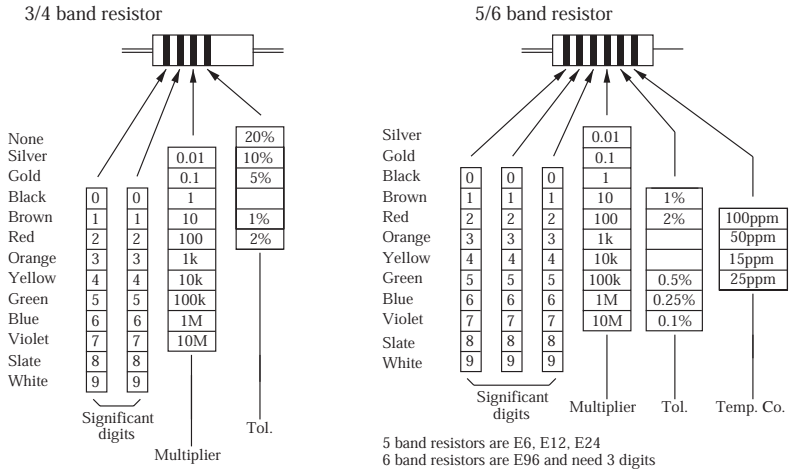


Fig. A.2 *Resistor colour codes*

Six band resistors

The first three bands denote the significant digits, and the fourth band is the multiplier.

The fifth band is the tolerance, note that six band resistors imply greater precision and so 5%, or worse, tolerance will not be seen.

Examples

Yellow, violet, yellow, red = 470 kΩ 2%
Yellow, violet, black, orange, brown, red = 470 kΩ 1% 50 ppm

Red, red, red, red = 2.2 kΩ 2%
Red, red, black, brown, brown, red = 2.2 kΩ 1% 50 ppm

Brown, black, black, red = 10 Ω 2%
Brown, black, black, gold, brown, red = 10 Ω 1% 50 ppm

Note that because the six band resistors have an extra significant digit, their multiplier is always one level lower than for the same value in a four band component.

Sometimes it can be difficult to decide which end of the resistor is which, and the value makes sense either way round:

Brown, orange, yellow, red = 130 kΩ 2%, but read the other way round = 24 kΩ 1%

If in doubt, measure the resistor with a digital multimeter; it is far easier to change the component now, than when it has been soldered into place.

Sallen & Key filters for active crossovers

In 1955, Sallen & Key[5] wrote a seminal paper on *RC* active filter design, and all component notation and page numbers hereon refer to this paper.

Cathode followers can be used to implement Sallen & Key filters, but they are by no means perfect. A low-pass filter has a pair of series resistors (R_1, R_2) at the input, and a capacitor is connected from the centre tap of these to the output of the cathode follower. Remembering that for the cathode follower, $r_{out} \approx 1/gm$, then typically $r_{out} \approx 500\,\Omega$, if $gm = 2$ mA/V. The output resistance forms a potential divider in conjunction with the feedback resistor or capacitor, which determines the maximum attenuation possible from the input *RC* pair, so a high *gm* valve maximizes stop band attenuation, which is $\approx R_1/R_{out}$.

As an example, a 12 dB/octave $f_{-3\,dB} = 860$ Hz 6J5 section was computer simulated as a possible contender for an active crossover. Practical values for the circuit resulted in a response falling to −54 dB at 18 kHz, at which point the response began to rise again at 6 dB/octave. Fortunately, two sections in series were proposed, and deviations at −108 dB almost out of the audio band were not felt to be very significant.

Because Sallen & Key filters rely on positive feedback to operate, they considerably worsen the distortion of their gain stage,[6] so it is crucial to minimize the distortion of the amplifier before it is enveloped by the filter. Clearly, the valve and operating point should be chosen carefully, but a slightly less obvious approach is to use a high μ valve which allows plenty of (distortion reducing) negative feedback before the positive feedback is applied, and further lower distortion with a constant current load, perhaps an EF184 pentode. At a typical operating point of the EC91 single triode, $\mu = 80$ and $gm = 5.7$ mA/V, making this an excellent theoretical choice, although the intrinsic linearity of the *N7/*SN7 family proved to be superior in the final analysis. See Fig. A.3.

Using the Sallen & Key paper

Unfortunately, the paper is somewhat tortuous, with much important information buried in obscure footnotes. The following example aims to guide you through the maze.

1 Determine the common cathode gain of the valve under chosen operating conditions, this is A_0. If we use a valve with a constant current load, then its gain is simply μ.

2 A cathode follower has 100% negative feedback, so the gain of the cathode follower is:

$$K = \frac{A_0}{A_0 + 1}$$

If we choose a 6J5 as the audio device, with a constant current load:

$$K = \frac{20}{21} = 0.9524$$

3 24 dB/octave Linkwitz–Riley[7] crossovers are made by cascading a pair of 12 dB/octave Butterworth filters, for which:

$$Q = \frac{1}{\sqrt{2}}$$

Alternatively, we can specify d:

$$d = \frac{1}{Q} = \sqrt{2}$$

4 It is easy to change resistor values, but capacitors are only available in a restricted range, so the two capacitor values should be equal:

$$\gamma = \frac{C_1}{C_2} = 1$$

5 The high-pass filter (which is the more awkward one to design) has a transfer function:

$$\frac{hs^2}{s^2 + ds + 1}$$

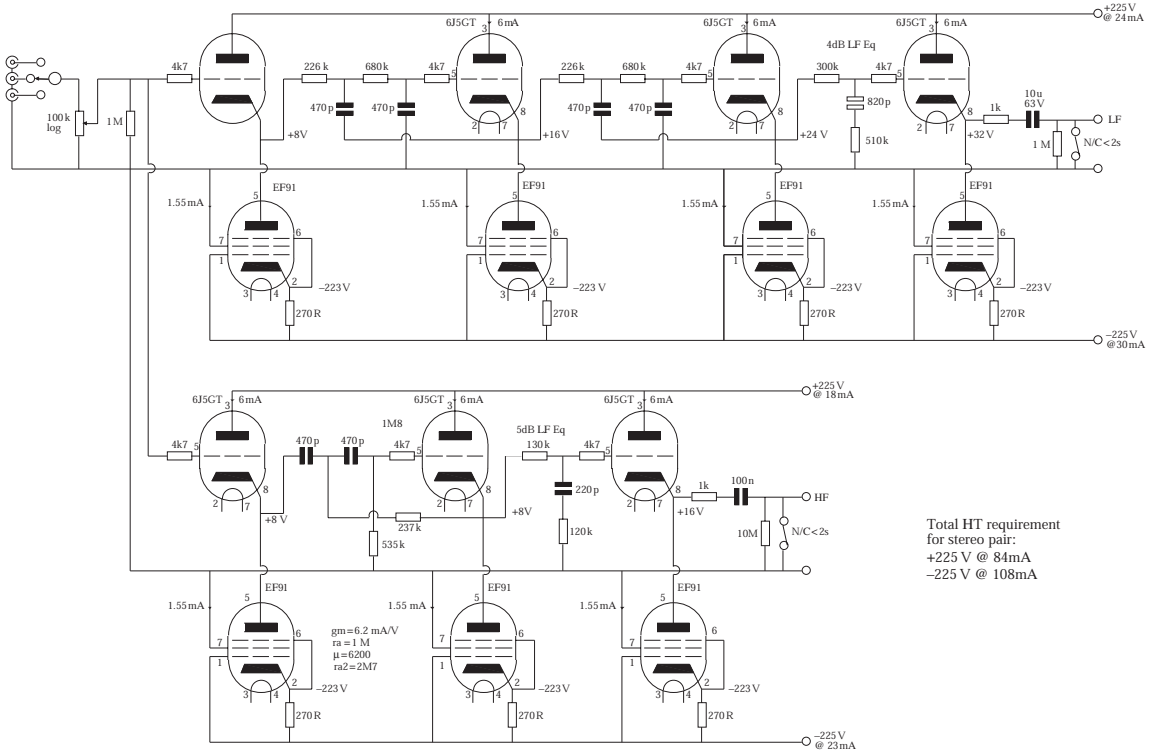


Fig. A.3 *DC coupled active crossover using symmetrical supplies*

We know that we will use a cathode follower (which is non-inverting), so we look at diagram 3 from the table on p. 76.

R_2 may well need to be a potential divider chain, such as on p. 75, which means that we are restricted in our choice of R_2, and because:

$$\tau_2 = R_2 C_2$$

Our choice of τ_2 has now been restricted.
(Alternatively, we can use a split rail supply as in Fig. A.3, or a THINGY to provide the bias.)

6 We now know that both γ and τ_2 are restricted, and we should therefore refer to Formulas Group III. This table contains general purpose formulae, which are explained by their footnote. We want to find τ_2, so we look to column b:

$$\frac{d}{2(1-K)} \cdot \left(1 \pm \sqrt{1 - \frac{4(1+x)(1-K)}{d^2}}\right)$$

We substitute our values: $\gamma = x = 1$, $K = 1$, $D = \sqrt{2}$

$$\tau_2 = \frac{\sqrt{2}}{2(1-0.9524)} \cdot \left(1 \pm \sqrt{1 - \frac{4(1+1)(1-0.9524)}{2}}\right)$$

The lower value of τ_2 is more stable (as explained on p. 81), so we only evaluate the equation for the case where we subtract the large square root term. We are calculating a time constant, so it must have units (seconds), giving:

$$\tau_2 = 1.493 \text{ s}$$

7 Referring back to the diagram on p. 76 to find τ_1, we see that:

$$\tau_1 \tau_2 = 1, \text{or}, \tau_1 = \frac{1}{\tau_2} = \frac{1}{1.493} = 0.6698$$

8 The filter designs have all been designed for a frequency (ω) of 1 radian/second, a procedure known as normalization. We want to be able to scale to our chosen frequency f (in Hz), so:

$$CR_2 = \frac{\tau_2}{2\pi f},\ CR_1 = \frac{\tau_1}{2\pi f}$$

We can now choose practical resistor and capacitor values for our chosen frequency.

9 (a) Remember that a high-pass filter biased by a potential divider has a value for R_2 that is the Thévenin resistance seen looking back into the divider chain (the two resistors in parallel). We choose resistor values to set our DC bias voltage first, then see if they produce a convenient capacitor value. Since the exact bias potential is unlikely to be critical, we can adjust the resistor values until a convenient capacitor value emerges.

(b) Having set the value of C, we can now calculate the value of R_1.

To find values for the low-pass filter, we repeat steps (6) to (9) using diagram 1 on p. 76, again setting $\gamma = 1$, so that we can use the same value of capacitor as used in the high-pass section. Not only is it cheaper to buy many capacitors of the same value, but it is much easier to change the frequency of the crossover later by changing standard capacitor values than by changing (and recalculating) non-standard resistor values.

UK sources of components

It used to be very difficult to obtain components for valve amplifiers, but there is now a plethora of vendors eager for your money. Valve suppliers can be found on the Internet simply by typing the valve type into a search engine. Alternatively, they can be found in Hi-Fi magazines – particularly those that tend to have constructional articles. Standard components can be bought from the following suppliers, and all stock a very limited range of valves.

Maplin Electronics

See adverts in electronics magazines for shop addresses.

RS Components

RS is a trade concern, but its retail arm 'Electromail' stocks all of the RS range. The RS CD-ROM is currently quite useful.

Farnell Electronic Components

Nominally a trade distributor, but FEC accepts orders (minimum order currently £30) placed on a personal credit card, so all you need is a current (paper) catalogue. The FEC CD-ROM is currently unfriendly.

Bargains and dealing directly

New parts

The previously mentioned companies distribute general electronic components, but do not necessarily stock specialist audio components such as Vishay resistors, etc. It is well worth shopping around for specialist components, as some stockists have imaginative pricing policies.

If you and your friends are able to club together to generate a large order it can be worth approaching manufacturers directly, after all, the worst they can do is to laugh at you. Components such as transformers and large film capacitors can often be made specially to order, and the author has had a number of transformers made by Sowter Transformers. If you choose to follow this course, specify as completely as possible what you need, and remember that every additional complication adds to the finished price.

Companies at Hi-Fi shows often give a 'show' discount on their goods, so it may be worth timing your order to coincide with a show. They may even be prepared to negotiate a further discount on the last day of the show.

New/second-hand parts

Electronics surplus shops are excellent places for picking up bargains, provided that you know what to look for. It is a good idea to take a digital multimeter, tape measure and calculator with you.

Amateur radio and vintage audio fairs are often fruitful sources of components, but bear in mind that most of the stands are occupied by traders who circulate from one fair to another, so they are unlikely to offer many goods at 'giveaway' prices. Nevertheless, bargains can be had, and otherwise awkward components can be found. In addition to components, there are usually many larger items on sale – the author was delighted to pick up a scruffy Garrard 401 for £30, but there are also boat anchors. ('Boat anchor' is the delightful amateur radio term for equipment whose main attribute is mass.)

When large companies close down a site, they often auction equipment, and this *can* be a source of carefully maintained test equipment. It is your responsibility to check the condition of what you buy, and you have no comeback afterwards. Don't get carried away at the auction, and remember that tax will be added to your bid price.

If you buy second-hand equipment from private advertisements, remember to add the cost of your return journey to inspect the goods, plus whatever cost is required to refurbish.

If you are not careful, Internet purchases can cost more than you thought. Carriage is always extra. If it is a foreign purchase, once it arrives in your country, it will be liable for import duty. This all adds costs, so do your sums before bidding. Nevertheless, the Internet is an extremely useful worldwide source for otherwise unobtainable items.

Even within your own country, second-hand goods can be expensive, because some sellers have wildly inflated ideas about the value of their goods, so bargains are often best found through your circle of friends, who are aware of your hobby.

References

1 Acoustical Manufacturing Company (1953–1967) *QC11 Owner's Manual*.

2 Lipshitz, Stanley P. (1977) On RIAA Equalisation Networks, *JAES*. June.

3 Langford-Smith (1953) *Radio Designers Handbook*, 4th ed. Iliffe, p. 730. Reprinted (1999).

4 Digest of Hi-Fi. Part Four: Record Replay (1990) *Hi-Fi Answers*, August.

5 Sallen & Key (1955) A Practical Method of Designing RC Active Filters, *IRE Transactions*, March.

6 Billiam, Peter J. (1978) Harmonic Distortion in a Class of Linear Active Filter Networks, *JAES*, June, Vol. 26, No. 6.

7 Linkwitz, Siegfried (1976) Active Crossover Networks for Non-coincident Drivers, *JAES*, Jan./Feb., Vol. 24, No. 1.

Index

AC loadline *see* Loadline
AC operating point *see* Operating point
Active crossovers, 457
  Sallen & Key filters, 596–600
Active devices, 44–5
Active loads, cathode follower, 109–10
Air dielectric, metal plate capacitors, 215–16
Air-cored inductors, 230–3
Alternating current (AC), 23–4
Ampere, definition, 3
Amplification factor μ, 72–3
Amplifiers:
  common collector amplifier, 52–3
  common emitter amplifier, 49–51
  see also B-follower; Common cathode triode amplifier; Differential pair; μ-follower; Operational amplifiers (op-amps); Power amplifiers; Pre-amplifiers; Push-pull amplifiers; Shunt Regulated Push-Pull (SRPP) amplifier; White cathode follower
Amplitude, 25
Analogue cassette tape output, 555
Analogue microgrove equalization, 590–1
Anode characteristics, common cathode triode amplifier, 64–7
Anode load/resistance, 70–1, 74, 379
Anodes:
  about anodes, 276
  characteristics, common cathode triode amplifier, 64–7
  heat dissipation, 276–8
  Secondary Emission Ratio (SER), 278–9
Audio transformers *see* Input transformers; Output transformers; Transformers

Balanced working:
  cathode followers, 552–3
  grid leak arrangement, 551
  hum loops, 554
  input stage, 548–51
  interconnects, 554
  line stage and volume control, 553
  Miller capacitance reduction, 551, 552
  pick-up arm wiring, 546–7
  second stage, 551–2
Band gap devices, 47–8
Bandwidth, noise bandwidth, 44
Batteries, 5
  rechargeable battery cathode bias, 174–5
Beam tetrodes, 88–9
  screen grids, 276
Beam triodes, 85–6
B-follower, 125–7
Bevois Valley amplifier, 445–55
  about this amplifier, 445–6
  author's prototypes, 453–5
  cathode bias/feedback resistor, 449–53
  DC conditions, 446–9
BH curves, 230, 382
Bias:
  bias current, 62
  see also Cathode resistor bias; DC bias; Operating point
Bifilar transformer windings, 238
Binary number system, 153–4
Bipolar Junction Transistors (BJTs):
  about BJTs, 48–9, 54
  common emitter amplifier, 49–51
  current gain, 48, 54
  Darlington pair, 53–4
  Early voltage, 51
  emitter follower, 52–3
  input and output resistances, 51–2
  transconductance (gm), 48
Blackman-Harris window, 156
Blocking, 190–1
  Crystal Palace amplifier, 466, 483
Blumlein ultralinear power output, 389
Boltzmann's constant, 44

Bridge rectifiers, 293–4
Bypass capacitors, 228–9
Bypass/cathode decoupling capacitors, 78–80

C core transformers, 236–8
Cable capacitance, 492
Capacitance, units for, 22
Capacitors:
about capacitors, 28–9, 211–12
air dielectric, metal plate, 215–16
bypassing techniques, 228–9
can potentials, 353
ceramic, 220–1
dielectrics, 213–14
effective series resistance (ESR), 214, 219, 307–8
electrolytic, 221–4
equivalent circuit, 214
film/foil, 216–19
heating effects, 226
leakage/insulation resistance, 214, 226
maximum ripple current, 214
metallized paper, 220
metallized plastic film, 219
microphony, 226–7
output coupling, 82–3
parallel plate, 212–13
polystyrene, 218–19
for power supplies, 306–10, 352–3
relative permittivity, 213
silver mica, 220
tantalum bead, 225
tantalum electrolytic, 225
tolerance, 225
types summary, 214–15
undischarged capacitor hazard, 353
units, 28
voltage rating, 225
Carbonization of valve envelopes, 180–1
Cascode:
about cascode, 94–5
with Crystal Palace amplifier, 473–6
design, 96–100
gain, 97–8
Hedge cascode differential pair, 458–60
semiconductor constant current sink, 134–5
Cathode:
bright emitters, 259
direct vs indirect heating, 263–5
dull emitters, 260
electromagnetic problem, 264–5
electrostatic problem, 263–4
heater tests, 267–71
heater voltage and current, 269–71
heater/cathode insulation, 265–6
heaters and their supplies, 267–71
hum problems, 264–5
Miller-Larson effect, 262–3
off/on cycle problems, 262–3
oxide-coated, 260
poisoning, 261
Stefan's law, 259
stripping, 261, 273
temperature considerations, 266–7
thermal problem, 263
thoriated tungsten, 259, 261–3
Cathode coupled phase splitter, 398–401
Cathode followers:
about Cathode followers, 100
with active load, 109–10
analysis/design, 100–4
with balanced working, 552–3
distortion from parameter restriction, 166–8
input capacitance, 103
linearity, 103–4
White cathode follower, 110–15
see also μ-follower
Cathode resistor bias:
about cathode bias, 75–8
for Bevois Valley amplifier, 449–53
bypass/decoupling capacitors, 78–80, 428–9
distortion problems, 172–3
effect on AC conditions, 77
on pre-amplifiers, 495
Scrapbox Challenge amplifier, 427–9
see also DC bias
CDs:
Equalization, 591–2
player signal levels, 555, 556

Ceramic capacitors, 220–1
Charge, 2–3
Choke, for Scrapbox Challenge amplifier power supply, 431–2
Choke input power supply, 310–15
Circuits, 5
  see also Equivalent circuits
Class *1 amplifiers, 384–5
Class *2 amplifiers, 385–6
Class A amplifiers, 383–4
  grid current problems, 164–5
  single-ended output stage, 378–80
Class B amplifiers, 384
Class C amplifiers, 384
Cockcroft-Walton voltage multipliers, 326–8
Coherent signals, and noise, 43
Common cathode triode amplifier, 64–86
  anode characteristics, 64–7
  anode dissipation, maximum, 68
  anode resistance, 70–1
  cathode bias, 75–8
  cathode current, maximum, 68
  cathode decoupling capacitor, 78–80
  constant current sink, 105–9
  cut-off, 67
  dynamic/AC conditions/parameters, 69, 72–5
  gain, 66–7, 73
  grid leak resistors, 80–2
  linearity, 69
  loadline, 65–6
  maximum undistorted voltage swing, 70
  Miller capacitance, 83–4
  operating point, 67–72
  output coupling capacitor, 82–3
  output resistance, 70, 85
  positive grid current, 68
  power supply, 65
  as pre-amplifier stage, 496–8
  space-charge limited condition, 64
  voltage amplification, 69–70
  voltage swing, maximum, 69–70
  see also Dynamic/AC parameters
Common collector amplifier, 52–3
Common emitter amplifier, 49–51
  stabilized, 50
Common-mode noise on LT supplies, 339–42
Common-Mode Rejection Ratio (CMRR), differential pairs, 130–2
Components:
  new parts, 601
  new/second-hand parts, 601–2
  UK sources, 600–1
Computer-Aided Design (CAD) AC analysis, EC8010 RIAA stage, 565–6
Computers:
  analogue computers, 57–8
  digital computers, 57–8
Concertina phase splitter, 403–7
  gain, 404
  output resistance, 405–7
Conductance, mutual, 74–5
Constant current sinks *see* Sinks, constant current
Contact potential, 161
Control grids:
  about control grids, 272–3
  and cathode stripping, 273
  frame-grid valves, 274
  grid current, 273
  grid emission, 273–4
  Richardson/Dushmann equation, 274, 584
  screen grids, 276
  thermal runaway, 273
  variable μ grids, 275–6
Copper losses, transformers, 239
Coulomb, definition, 3
Coupling:
  coupling capacitors, 67, 82–3
  DC coupling and Level shifting, 192–4
  transformer interstage, 191–2
  see also Interstage coupling
CR circuits, 39–41
Crossovers, active Sallen & Key filters, 596–600
Crystal Palace amplifier, 460–83
  334Z constant current sink, 476–7
  blocking, 466, 483
  cascode constant current sink, 473–6
  chassis, 480

Crystal Palace amplifier, (C*ontinued*)
constant current sink, 3347 and thermal stability, 476–7
constant current sinks and power supplies, 466–8
driver requirements, 463–6
global negative feedback, 483
HF stability, 477–80
HT regulators, 480
output stage, 469–72
output valves, 460–1
overload recovery and blocking, 466
power supplies, 468–9, 470–2, 480–2
thermal stability, 476–7
valve matching, 472–3
Current:
alternating, 23–4
definition, 3
direct, 23
and electron flow, 45
Current sinks *see* Sinks
Cut-off, common cathode triode amplifier, 67
Cut-off frequency, filters, 32

DAB, output levels, 555
Darlington pair, 53–4
DAT, output levels, 555
Daughter of Beast, 483–7
DC bias:
Bevois Valley amplifier, 449–53
cathode resistor bias, 173–4
common cathode triode amplifier, 67, 75–7
constant current sink bias, 178
diode cathode bias, 175–8
grid bias voltage supply, 173–4
rechargeable battery cathode bias, 174–5
Scrapbox Challenge amplifier, 427–8
see also Cathode resistor bias
DC coupled Class A electromagnetic headphone amplifier, 194–7
DC coupling and Level shifting, 192–4
DC (Direct Current), 23
DC offsets, 62
DC operating point *see* Operating point
Decibel (dB), 22–3

Dielectrics:
dielectric absorption, 217
film/foil capacitors, 218
strength and loss, 213–14
Differential pairs:
basic principle, 127–9
Common-Mode Rejection Ratio (CMRR), 130–2
differential triode pair, 395
distortion cancellation effects, 169–70
gain, 129
output resistance, 129–30
Power Supply Rejection Ratio (PSRR), 132–3
Digital measurement:
analogue to digital conversion, 151
anti-aliasing filters, 152
Fast Fourier Transform (FFT), 154–5, 156
Moiré patterns, 152
number systems, 153–4
Nyquist criterion, 152
periodicity assumption, 155
precision, 154
quantization, 153
sampling, 152
scaling, 152–3
windowing, 155–6
Diode cathode bias, 175–8
Diodes:
band gap devices, 47–8
diode drop, 46
forward drop, 47
Light Emitting Diodes (LEDs), 47
silicon, 45–6
types of, 47
voltage references, 46–8
Zener diodes, 47
Direct current (DC), 23
Distortion:
and AC operating point, 157–60
and battery cathode bias, 174–5
cathode resistor bias, 173–4
and Class A2 amplifiers, 164–5
and DC operating point, 161
differential pair cancellation, 169–70
and diode cathode bias, 175–8

and feedback, 55
from grid current at contact potential, 161
from grid current and volume controls, 161–4
and grid bias voltage, 173–4
linear distortion, 143
non-linear distortion, 144
and parameter restriction, 165–8
push-pull cancellation, 169
reduction by cancellation, 169–70
single ended output stage, 380
slewing distortion, 43
and timing errors, 143
valve distortion, 147
weighted results, 188–9
see also DC bias; Interstage coupling; Noise; Valve selection and testing
Distortion measurement:
about measuring distortion, 145–6
alternative rectifiers, 150
author's measurement method, 156–7
choice of measurement, 146–7
differential pair distortion cancellation, 170–1
distortion residual, 148–50
harmonic distortion, 147–8
intermodulation distortion, 145
μ-follower linearity test circuit, 159
noise and THD, 150–1
non-linear distortion, 144–5
spectrum analysers, 151
test equipment limitations, 146
Total Harmonic Distortion (THD), 145, 148
weighting harmonics, 147–9
wow and flutter, 146
see also Digital measurement
Dither, digital systems, 487
Driver stage, 394–6
for Scrapbox Challenge amplifier, 435–9
DVD output level, 555
Dynamic/AC parameters:
about dynamic characteristics, 72
amplification factor μ, 72–3
anode resistance, 70–1, 74, 379
mutual conductance (gm), 74–5
E88CC vs ECC88 heater current, 584
Early voltage, bipolar junction transistors, 51
Ebers/Moll equation, 46–8, 49
EC8010 RIAA stage, 557–63
75 μs problem, 565
75 μs/3.18μs manipulation, 566–7
3180 μs pairing manipulation, 566
Computer-Aided Design (CAD) solution, 565–6
direct measurement problems, 570–1
grid current distortion, 564
input stage, 557–60
input transformer, 560–1
line stage, 572–7
Miller capacitance problems, 564–5
practical considerations, 570–2, 577
production tolerances and component selection, 571
quiescent current, 572–4
RIAA equalization design, 563–9
second stage, 561–2
output stage, 562–3
valve choice, 563, 574–6
valve tolerances, 571–2
see also RIAA (Recording Industry Association of America) record equalization
Effective Series Resistance (ESR), capacitors, 214, 219, 307–8
Electro-Motive Force (EMF), 4, 9
Electrolytic capacitors:
about electrolytic capacitors, 221
bipolar, 224
construction, 221–3
electrolyte evaporation, 223
life, 223–4
limitations, 223
ratings, 224
tantalum, 225
Electrons, 2–3
charge/mass ratio, 258
electron flow, 45
electron velocity, 257–8
Electrostatic hysteresis, 217
Emission, 257
Emitter follower, 52–3

Equalization:
analogue disc standards, 589–91
analogue microgrove, 590–1
CDs, 591–2
networks, 33–4
NICAM 728 recording, 592–3
see also EC8010 RIAA stage; RIAA
Equivalent circuits, 16
Norton, 20–1
Thévenin, 16–20
Excess/flicker noise, 44
Exponential waveforms, 39–42

Farad (F), 28
Farnell Electronic Components, 601
Fast Fourier Transform (FFT), 154–5, 156
Feedback *see* Negative feedback; Positive feedback
Feedback equation, 54–6
limitations, 56
Film/foil capacitors, 216–19
dielectrics for, 218
Filtering for power supplies:
high frequency, 322–5
low frequency, 320–2
Filters, 30–3
cut-off frequency, 32
high-pass, 30–1
LC circuits, 40, 320–5
low-pass, 32
and phase, 32
RFI, 355–7
Sallen & Key for active crossovers, 596–600
Flicker/excess noise, 44, 534
Floating paraphase/see-saw phase splitter, 401–3
FM analogue radio output, 556
Fourier series, 38
Frame-grid valves, 274
Frequency and radians, 25
Futterman Output Transformer-Less (OTL) amplifiers, 393

Gain, 73
Bipolar Junction Transistors (BJTs), 48, 54
cascode, 97–8
common cathode triode amplifier, 66–7, 73
concertina phase splitter, 404
differential pairs, 129
pentodes, 92
pre-amplifiers, 498–500
RIAA equlization gain equation, 523–5
Getters, 281–3
Glass Reinforced Plastic (GRP) boards, 289
Grain Oriented Silicon Steel (GOSS), 237
Grid bias *see* DC bias
Grid current, 273
at contact potential, 161
and Class A2 amplifiers, 164–5
distortion in EC8010 RIAA stage, 564
and volume controls, 161–4
Grid emission, 273–4
Grid leak resistors, 76, 80–2
Grid stoppers, 411
Grids *see* Control grids
Guided-grid triodes, 85–6
Guitar amplifiers, transformers with, 254

Headphone amplifier, DC coupled Class A, 194–7
Headphone direct drive amplifier, 483–7
heat and cooling, 488
power supply hum, 485–7
Heat/cooling:
anodes, 276–8
capacitors, 226
headphone direct drive amplifier, 488
resistors, 203–4, 211
Heater supplies *see* Low Tension (LT) heater supplies
Hedge cascode differential pair, 458–60
Henry (H), 28
High Tension (HT) supplies:
capacitor can potentials, 353
capacitor ratings, 352–3
centre-tapped and bridge, 293–4
choke current rating, 313–15
choke input minimum load current, 312–13

choke input ripple voltage, 310–11
for common cathode triode amplifier, 65
Crystal Palace amplifier, 468–9
current spikes, 315–18
errors and failures, 370–1
hard valve rectifiers, 294–7
HT delay circuit, 375
HT voltage choice, 351–2
LC filters at high frequencies, 322–5
LC filters at low frequencies, 320–2
mercury vapour rectifiers, 298
output volts and ripple, 301–2
a practical design, 351–65
Radio Frequency Interference (RFI), 355–7
reservoir capacitors, 306–10, 318–20
RF noise problems, 298–9
ripple current considerations, 302–5
ripple voltage with choke input, 310–11
ripple voltage with single reservoir, 299–301
for Scrapbox Challenge amplifier, 429–35
sectioned RC filters, 325–6
silicon vs hard valve rectifiers, 294–7
snubbers, 316–18
standby/operating mode, 369–70
switch-on current surge, 354–5
transformer core saturation, 305–6
transformer current rating, 315
transformer/reservoir capacitor combination, 306–10
undischarged capacitor hazard, 353
voltage multipliers, 326–8
wide-band filtering, 324–5
zero voltage switching relay, 365
see also Rectifiers; Voltage regulators
High-pass filters, 30–1
HT transformers *see* High Tension (HT) supplies; Transformers
Hum:
and balanced working, 554
cancellation with valve voltage regulators, 345
cathode problems, 263–5
with headphone amplifier, 485–7
with unbalanced input stages, 554–5
Hysteresis, transformers, 235–6

Impedance, 30
Incoherent/random noise, 44
Inductance:
transformer leakage, 236
units, 28
wirewound resistors, 208–10
Inductors:
about inductors, 28–30, 229–30
air-cored, 230–3
BH curves, 230
gapped cores, 233
relative permeability, 230
self-capacitance, 233–4
Input referred noise, 533
Input resistance/impedance, 16
and feedback, 55, 56–7
μ-follower, 117
Input selectors, 513–16
mercury wetted relays, 515–16
switch quality, 515–16
Input stage to power amplifiers, 407–8
global feedback considerations, 407
Mullard 5–20 amplifier, 415
Williamson amplifier, 412–14
Input transformers loading, 245–7
to EC8010 RIAA stage, 560–1
and Zobel networks, 245–7
Instability:
motorboating/LF instability, 410–11
parasitic oscillation, 411–12
squegging, 411
Insulation/leakage resistance, capacitors, 214
Integrating op-amps, 61–2
saturation, 62
Intermodulation distortion, 145
Interstage coupling, 189–200
blocking problems, 190–1
DC coupled Class A electromagnetic headphone amplifier, 194–7
level shifting and DC coupling, 192–4
Norton level shifter, 193–4, 198–200
transformer coupling, 191–2
Inverting op-amps, 58–9
Ionization noise, 280–1

Johnson/thermal/white noise, 43
Joule, definition, 7

Kirchoff's laws, 8–10
current law, 8–9, 11
voltage law, 9–10

LC circuits/filters, 40
with power supplies, 320–5
Leak amplifiers, 241
Leakage inductance, 236
Leakage/insulation resistance, capacitors, 214
Level shifting and DC coupling, 192–4
LF instability/motorboating, 410–11
Light Emitting Diodes (LEDs), 47
Light-sensitive resistors, 511–12
Lightbulbs, 5, 6, 8
Line stage *see* Pre-amplifiers
Lipshitz networks, RIAA record equalization, 525–6
LM334Z IC constant current sink, 140
Loadline:
μ-follower, 115, 118–19
common cathode triode amplifier, 65–6
single ended output stage, 379
Logarithmic law, for volume controls, 501–3
Loudspeakers:
efficiency, 456
horn, 381
power compression, 456
reflex, 381
Low Tension (LT) heater supplies:
for a better power supply, 367–8
common-mode noise, 339–42
current regulator design, 368–9
design considerations, 357–9
elevated voltage references, 362–5
for HT supply rectifiers, 355
references for elevated heater supplies, 362–5
supply arrangements, 267–71
transformers, 371–2
Low-pass filters, 32
LT transformers *see* Low Tension (LT) heater supplies; Transformers

Magnetic components *see* Inductors; Transformers
Magnetostriction, transformers, 240
Magnification factor (Q), 34–5
and parasitic oscillation, 411
Maida regulator, 480
Maplin Electronics, 600
Mathematical symbols, 1–2
MD, output level, 555
MD audio output, 555
Mercury vapour rectifiers, 298
Mercury wetted relays, 515–16
Metal film resistors:
manufacturing process, 204–6
max voltage rating, 206–7
modulation noise, 207
soldering, 207
tracking problems, 206–7
Metallized paper capacitors, 220
Metallized plastic film capacitors, 219
Mica capacitors, 220
Microgroove equalization, 590–1
Microphony, with capacitors, 226–7
Miller capacitance:
with balanced working, 551, 552
with EC8010 RIAA stage, 564–5
tetrodes, 88
triodes, 83–4
Miller-Larson effect, 262–3
Modulation noise, metal film resistors, 207
Moiré patterns, 152
Motorboating/LF instability, 410–11
μ-follower, 115–18
AC loadline, 115, 118–19
input resistance, 117
limitations, 119–20
Power Supply Rejection Ratio (PSRR), 118
upper valve choice, 119
Mullard 5–20 power amplifier, 415–20
Multifilar transformer windings, 239
Multipliers and units, 21–2
Mutual conductance (gm), 74–5

Negative feedback, 55
on Bevois Valley amplifier, 449–53
DC coupled headphone amplifier, 196
and distortion, 55
feedback equation, 54–6

feedback factor, 57
feedback fraction, 55
global feedback, 407
and input/output resistance/impedance, 55, 56–7
LF instability/motorboating, 410–11
negative feedback, 55, 56–7
and output transformers, 240–1
and positive feedback, 55–7
with pre-amplifiers, 499–500
scrapbox Challenge amplifier, 439
slugging the dominant pole, 409–10
and stability, 408–12
step networks, 410
terminology, 56–7
see also Instability
NICAM 728 TV audio:
equalization, 592–3
output, 556
Noise:
Boltzmann's constant, 44
flicker/excess noise, 44, 534
input inferred noise, 533
noise bandwidth, 44
noise powers, 44
noise and THD, 150–1
in pre-amplifiers, 498–500, 528–37
random/incoherent, 43–4
resistor excess noise, 532
resistor thermal noise, 531–3
with RIAA equalization, 536
valve noise, 533–4, 536–7
white/Johnson/thermal, 43, 533
wirewound resistors, 208–10
Nomenclature convention, 187–8
Non-inverting amplifier, 60–1
Norton equivalent circuit, 20–1
Norton level shifter, 193–4, 198–200
Nyquist criterion, 152

Octaves, 31
Offset current and voltage, 62
Ohm's Law, 5–6, 7, 12
Operating point:
AC point and distortion, 157–60
common cathode triode amplifier, 67–72
DC point and distortion, 161–5
Scrapbox Challenge amplifier, 425–7, 435–9
single-ended output stage, 379
Operational amplifiers (op-amps):
about op-amps, 57–8
bias current, 62
integrators, 61–2
inverters, 58–9
non-inverting amplifier, 60–1
offset current, 62
offset voltage, 62
virtual earth adder, 59–60
voltage follower, 60–1
Oscillation, parasitic, 411–12
Output amplifiers *see* Power amplifiers; White cathode follower
Output coupling capacitors, 82–3
Output resistance/impedance, 16
common cathode triode amplifier, 70, 85
and feedback, 55, 56–7
loudspeaker considerations, 381
Output Transformer-Less (OTL) amplifiers, 393
Output transformers, 240–2
with Blumlein/ultralinear output stage, 389
cathode follower driven, 389–90
and feedback, 240–1
imperfections, 381–3
for push-pull amplifiers, 386–8
Scrapbox Challenge amplifier, 427
see also Transformers
Output valves:
13E1 for Crystal Palace amplifier, 461–3
6528 for Scrapbox Challenge amplifier, 424
EL34 for Mullard 5–20 amplifier, 415–17
KT66 for Quad II amplifier, 419
KT66 for Williamson amplifier, 414
single-ended Class A, 378–80
transformerless amplifiers, 393

Parallel output valves, 458
Parallel plate capacitors, 212–13

Parallel resonant circuit, 35–6
Parasitic oscillation, 411–12
Partition noise, 94
PCBs (Printed Circuit Boards), 204, 289
Peak (pk), definition, 24–5
Pentodes:
  about pentodes, 88
  constant current sink, 107–9
  curves for, 89–91
  distortion spectrum, 90
  EF86 small-signal, 91–4
  gain, 92
  screen grids, 276
  transconductance (gm), 93
Periodicity assumption, 155
Permeability, relative, 230
Permittivity, relative, 213
Phase, and filters, 32
Phase splitters:
  basic principles, 396–8
  cathode coupled, 398–401
  concertina phase splitter, 403–7
  differential/long-tailed pair, 398–9
  floating paraphase/see-saw splitter, 401–3
  Mullard 5–20 amplifier, 415
  Quad II power amplifier, 419–22
  transformer splitting, 397
Pick-up arms *see* Vinyl records
Pinstriping transformers, 382–3
Plastic film, foil plate capacitors, 216–19
Polystyrene capacitors, 218–19
Positive feedback, 55
  and input/output impedance, 56–7
Positive grid current, 68
Potential difference (pd), 4
Potential dividers, 15, 18
Potentiometers *see* Volume controls
Power, 6–8
Power amplifiers, 377–488
  about output stages, 377–8
  above 10 W, 455–6, 458–83
  Bevois Valley amplifier, 445–53
  cathode follower circuit, 389–92
  classes of, 383–6
  Crystal Palace amplifier, 460–83
  driver stage, 394–6
  for electrostatic headphones, 483–7
  equal anode/cathode loads, 390–2
  input characteristics, 494
  input stage, 407–8
  Mullard 5–20 amplifier, 415–19
  negative feedback and stability, 408–12
  output impedance problems, 380–1
  Output Transformer-Less (OTL) amplifiers, 393
  parallel output valves, 458
  phase splitters, 396–407
  Quad II amplifier, 419–23
  single-ended Class A, 378–80
  transistor driven, 456
  ultralinear (Blumlein), 389
  Williamson amplifier, 412–15
  see also Negative feedback; Output transformers; Phase splitters; Push-pull amplifiers; Scrapbox challenge
Power supplies:
  a better power supply, 365–75
  Crystal Palace amplifier, 468–9, 470–2, 480–2
  electrostatic headphone amplifier, 485–7
  switchers (switch mode), 292–3
  see also High Tension (HT) supplies; Low Tension (LT) heater supplies
Power Supply Rejection Ratio (PSRR):
  μ-follower, 118
  differential pairs, 132–3
  to individual stages, 348–51
Pre-amplifiers, 491–578
  about pre-amplifiers, 491
  author's prototype, 544–6
  basic compromises, 547
  common cathode triode amplifier design, 496–8
  flicker noise, 534
  frequency characteristics, 492–3
  gain considerations, 498–500
  hum problems, 554–5
  input selectors, 513–15
  input sensitivity, 496
  input stage, 528–37
  line stage, 491–500
  linearity and headroom, 492
  negative feedback, 499–500
  noise considerations, 498–500, 528–37

output characteristics, 492–6
potentiometers, 496
switched attenuator volume controls, 503–11
traditional line stage, 496
valve choice, 498, 529, 534–6
valve noise, 533–4
see also Balanced working; EC8010 RIAA stage; Noise; RIAA; Volume controls
Printed Circuit Boards (PCBs), 204, 289
Push-pull amplifiers:
above 10 W, 455–6
Bevois Valley amplifier, 445–53
Crystal Palace amplifier, 460–83
distortion cancellation in, 169
Hedge cascode differential pair, 458–60
Mullard 5–20 amplifier, 415–19
output stage, 386–8
Quad II amplifier, 419–23
Williamson amplifier, 412–15
see also Power amplifiers; Shunt Regulated Push-Pull (SRPP) amplifier

Q (magnification factor), 34–5
and parasitic oscillation, 411
QBASIC:
law faking resistor calculations, 502–3
RIAA circuit calculation, 540
switched attenuator design, 506–8
Quad II power amplifier, 419–23
Quantization, 153

Radians and frequency, 25
Radio Frequency Interference (RFI), 355–7
Random/incoherent noise, 43–4
Reactance, 29–30
Rechargeable battery cathode bias, 174–5
Recording Industry Association of America *see* RIAA
Rectifiers:
bridge, 293–4
hard valve types, 294–6
mercury vapour rectifiers, 298
for practical power supply, 360–2
for Scrapbox Challenge amplifier, 430
in series for HT supplies, 373
silicon vs hard valves, 294–6
Regulators *see* Voltage regulators
Relative permeability, 230
Relative permittivity, 213
Resistance:
definition, 5–6
product over sum equation, 11, 19
see also Input resistance/impedance; Output resistance/impedance
Resistors, 202–11
colour codes, 594–6
excess noise, 532
heating/temperature rise/power dissipation, 203–4, 211
metal film, 204–7
preferred values, 202–3
on printed circuit boards, 204
series/parallel networks, 10–14
standard values, 593–4
thermal noise, 531–3
tolerance considerations, 211
voltage rating, 211
wirewound, 207–10
Resonance, 34–6
RIAA (Recording Industry Association of America) record equalization, 516–28
about record equalization, 34, 516–17
arm and coil wiring resistance, 519–20
equalization gain equation, 523–5
equalizer design, 520–3
Lipshitz networks, 525–6
mechanical problem, 517–19
noise and distortion improvement, 536, 557–60
output levels, 556
practical implementation, 537–47
pre-amp location, 529
requirements, 520–3
split equalization, 526–8
see also EC8010 RIAA stage
Richardson/Dushmann equation, 274, 584
RMS (Root of the Mean of the Squares), 36–7
RS Components, 600

Sallen & Key filters for active crossover, 596–600
Sampling, analogue to digital conversion, 152
Sawtooth waveforms, 26
Scrapbox Challenge single ended amplifier, 424–44
  bias, 427–8, 438
  cathode bypass capacitor, 428–9
  designer's observations, 442–4
  driver stage, 435–9
  feedback, 439
  HT regulator, 432–5
  HT supplies, 429–35, 441
  listening tests, 442
  operating point, 425–7
  output class, 425
  output transformer, 427
  output valve, 424–5
  teething problems, 439–42
Screened-grid valves, 85
Secondary Emission Ratio (SER), 278–9
See-saw phase splitter, 401–3
Semiconductor constant current sinks, 133–5
Semiconductor devices/circuits:
  LM334Z IC constant current sink, 140
  N-type and P-type, 45–6
  useful transistor types, 138–40
  as valve active loads, 135–8
  see also Bipolar Junction Transistors (BJTs); Diodes; voltage references
Series resonant circuit, 35
78 records *see* Vinyl records
Shunt Regulated Push-Pull (SRPP) amplifier:
  about SRPP amplifiers, 120–3
  applications, 125
  principle of operation, 122–5
Silicon diodes, 45–6
Silvered mica capacitors, 220
Sine waves, 24–6
Single-ended Class A power amplifier, 378–80
Sinks, constant current:
  cascode current source, 134–5
  for common cathode amplifier, 105–9
  Crystal Palace amplifier, 466–8
  current, 104–5
  LM334Z IC, 140
  pentode as, 107–9
  semiconductor type, 133–5
Sinks, constant voltage, 104–5
Slugging the dominant pole, 409–10
Snubbers, power supplies, 316–18
Sockets *see* Valve sockets
Sources:
  current, 104–5
  voltage, 104–5
Space-charge limited condition, triodes, 64
Spectrum analysers, 151
Spreadsheets, volume control calculations, 509–10
Square waves:
  about square waves, 37–8
  and CR and LC networks, 38–9
  square wave sag, 585–7
Squegging, 411
Stability:
  and feedback, 408–12
  motorboating, 410–11
  and parasitic oscillation, 411–12
  slugging the dominant pole, 409–10
  squegging, 411
Standard component values, 593–4
Standby/operating mode, power supplies, 369–70
Stefan's law, 259
Step networks, 410
Step waveforms, 38–9
Switch-on current surge, 354–5
Switched attenuator volume controls:
  about switched attenuators, 503–5
  construction, 510–11
  design, 505–10
Switchers (switch mode) power supplies, 292–3
Switches, 5
Synthetic Resin Bonded Paper (SRBP) boards, 289

Tantalum electrolytic capacitors, 225
Tape output:
  1/4 inch analogue, 555
  analogue cassette, 555

Tertiary windings, transformers, 27
Tetrodes:
  about tetrodes, 86–8
  anode characteristics, 88
  Miller capacitance, 88
  screen grids, 276
  secondary emission kink, 88
  see also Beam tetrodes; Pentodes
Thermal runaway due to grid current, 273
Thermal/white/Johnson noise, 43
Thermionic emission, Richardson/Dushmann equation, 274, 584
Thermionic valves *see* Valves, thermionic
Thévenin equivalent circuit, 16–20
Thévenin resistance, 538, 542
Time constants, 33–4
Tolerances:
  and awkward values, 544
  EC8010 RIAA stage, 571–2
Toroidal transformers, 238
Total Harmonic Distortion (THD), 145, 148, 150–1
Transconductance (gm), 48
  pentodes, 93
Transformers:
  applications, 247–8
  audio, 253–4
  basic principles, 26–7
  BH curve, 382
  bifilar windings, 238
  C cores, 236–8
  centre-tapped, 373–5
  choosing, 248–9
  copper losses, 239
  core saturation, 305–6
  damage problems, 254–5
  DC magnetisation, 239
  eddy current losses, 235–6
  electrostatic screens, 240
  equivalent circuits, 242–4
  with guitars, 254
  hysteresis loss, 235–6
  identifying unknown transformers, 249–53
  interstage coupling, 191–2
  iron losses, 235–9
  for Leak amplifiers, 241
  leakage inductance, 236
  LT transformers, 371–2
  magnetostriction, 240
  models for, 242–5
  multifilar windings, 239
  perfect, 27
  pinstriping, 382–3
  for power supplies, 306–10, 315
  for Scrapbox Challenge amplifier HT supply, 431
  tertiary windings, 27
  toroidal arrangement, 238
  turns ratio, 27–8
  uses and abuses, 253–5
  see also Input transformers; Output transformers
Transient waveforms, 38–43
Transistor amplifiers:
  input resistance, 494
  input voltage requirements, 494
Transistors:
  power amplifiers, 456
  useful types, 138–40
  see also Bipolar Junction Transistors (BJTs); Semiconductor devices/circuits
Triodes:
  guided-grid/beam triodes, 85
  see also Common cathode triode amplifier
Turns ratio, transformers, 27–8
Twisted pair working, 546–7

Ultralinear (Blumlein) amplifier, 389
Units and multipliers, 21–2

Valve data:
  American consumer valve codes, 582–4
  books, 579
  E88CC vs ECC88 heater current, 584
  European Pro-Electron valve codes, 580–2
  Internet, 580
  pin connections, 584
  thermionic emission, 584

Valve noise, pre-amps, 533–4, 536–7
Valve selection and testing:
audio test interpretation, 184–7
audio test methods, 183
audio test results, 184
author's summary, 189
and carbonization of envelopes, 180–1
distortion weighted results, 188–9
EC8010 RIAA stage, 563
and electron deflection, 181
low distortion considerations, 179–80
low distortion test circuit, 182–3
low distortion testing, 181–2
medium μ valves, 187–8
nomenclature of book, 187
pre-amplifiers, 534–6
Valve sockets:
LoktalTM base, 287
pin plating, 288
pins and the envelope, 287–8
PTFE, ceramic and phenolic, 286–7
Valve voltage regulators *see* Voltage regulators
Valves, thermionic:
cooling methods, 284–6
electron charge/mass ratio, 258
electron velocity, 257–8
emission, 257
envelope temperature measurement, 284
frame-grid valves, 274
getters, 281–3
history, 255–7
ionization noise, 280–1
mica wafer problems, 282–4
vacuum considerations, 279–81
see also Anodes; Cathode; Common cathode triode amplifier; Control grids; Output valves; Pentodes; Tetrodes; Triodes; Valve data; Valve selection and testing; Valve sockets
VCR:
analogue, 556
NICAM TV sound, 555
Vinyl records:
cartridge output level, 556
equalization standards, 589–91
grove size, 587–8
pick-up arm mechanics, 588–9
pick-up arm wiring, 546–7
playing 78s, 587–9
speed problems, 587
see also EC8010 RIAA stage; RIAA
Virtual earths, 58–9
virtual earth adder, 59–60
Volt, definition, 4
Voltage follower op-amp, 60–1
Voltage multipliers, 326–8
Voltage references, 46–8
band gap devices, 47–8
slope resistance, 47
Zener diodes, 47
Voltage regulators:
317 IC voltage regulator, 336–9, 342–3
about regulators, 328–9
for a better power supply, 372–3
drop-out voltage, 331
fundamental series regulator, 329–31
hum cancellation, 345
increasing output current, 346–8
optimized regulator, 344–5
output inductance compensation, 334
for practical power supply, 360–2
for Scrapbox Challenge amplifier, 432–5
sound from, 351
speed-up capacitor, 332–4
two-transistor series regulator, 331–2
valve type, 343–8
variable bias voltage regulator, 335–6
Voltage sinks *see* Sinks
Volume controls, 496, 500–16
balanced, 512–13, 553
and distortion with grid current, 161–4
with light-sensitive resistors, 511–12
logarithmic law faking, 501–3
logarithmic law requirement, 501
QBASIC calculations, 502–3, 506–8
skelatel type, 501
spreadsheet calculations, 509–10
switched attenuator controls, 503–11

Watt, definition, 6–7
Wein bridge, 43

White cathode follower:
as an output stage, 113–15
analysis, 111–13
White/Johnson/thermal noise, 43
Williamson amplifier, 412–15
stability, 414
Windowing, analogue to digital conversion, 155–6
Wirewound resistors:
ageing, 208
inductive reactance, 208–10
manufacturing process, 207
noise, 208
Work, definition, 4
Wow and flutter, 146

Zener diodes, 47, 176
in voltage regulators, 335–6
Zobel networks, 245–7, 457–8

This page intentionally left blank